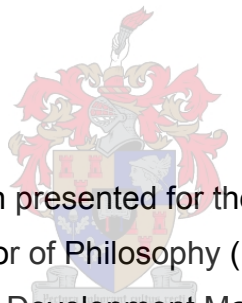
The potential value of the Realist Evaluation Method in programme impact evaluations in South Africa

Patience Nombeko Mbava



Dissertation presented for the degree of
Doctor of Philosophy (PhD)
(Public and Development Management)
in the Faculty of Economic and Management Sciences
at Stellenbosch University

Promoter: Dr Babette Rabie

December 2017

Declaration

By submitting this thesis electronically, I, Patience Nombeko Mbava, declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

P.N. Mbava

December 2017

Abstract

In the context of the development of a coherent country-driven monitoring and evaluation (M&E) system in the South African public sector underpinned by the policy outlining the Government-Wide Monitoring and Evaluation System (GWMES), this study explored the methodological approaches applied in impact evaluations. With the field of evaluation theory still booming internationally, there are no prescribed guidelines as to which impact evaluation method, employed in a specific context and under which conditions, will render the most useful findings for policy-makers. Therefore, clarifying a suitable methodological approach for impact evaluation of social programmes is critical in South Africa. Towards these aims, this study explored appropriate methodological approaches in informing better programme impact evaluations through the exploration of the potential value of Realist Evaluation Method (REM) on impact evaluations within the South African public sector.

The research had a three-pronged approach. First, a comprehensive literature review was conducted to gain an in-depth understanding of the REM through a detailed review and analysis of the related literature, as well as to assess the current trends in research and application of REM approach in conducting impact evaluations. Secondly, case study micro-analyses were completed to investigate the methodologies and approaches used in past programme impact evaluations as well as to establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects. Finally, in-depth interviews with policy decision-makers, commissioners and implementers of evaluations were completed to further ascertain the utility value of evaluation results as well as establish the applicability of the REM as a methodological approach in conducting programme impact evaluations in the South African public sector.

The overall findings provided evidence that initial well-defined and coherent programme theory as well as programme causality on commissioned impact evaluations are largely absent. The research also found limited contextual understanding of the programmes' intersection with the broader complex social system. This offered policy-makers limited insights in terms of understanding for whom a social programme will work most effectively or not and the reasons thereof. Therefore, this limited explanatory focus resulted in impact evaluations that had a 'black box' phenomenon, as the key change mechanism in programmes were unknown.

It was concluded that REM has a potential value to contribute to programme impact evaluations that offer new insights regarding what works, for whom, under what conditions and in what respects. However, there are potential constraints in its application and these should be well considered against the benefits that can be derived from such evaluations.

A key contribution of the study is an assessment model, applicable to the South African context, that could be applied to determine, from a Realist Evaluation lens, the value of an evaluation to different policy-makers' needs and can be applied to assess the limitations of other impact evaluation methods.

Key words:

impact evaluation

public sector,

monitoring and evaluation

realist evaluation method

Opsomming

In die konteks van die ontwikkeling van 'n samehangende landgedrewe monitering- en evaluering-stelsel (M&E) in die Suid-Afrikaanse openbare sektor, ondersteun deur die beleid wat die regeringswye monitering- en evalueringstelsel (GWMES) uitstip, het hierdie studie die metodologiese benaderings wat op impak-evaluerings toegepas word, ondersoek. Aangesien die studieveld oor evalueringsteorieë internasionaal steeds groei, is daar geen voorgeskrewe riglyne nie oor watter impak-evaluasie metode, in watter spesifieke konteks en onder watter omstandighede, die mees bruikbare bevindinge vir beleidmakers sal lewer. Daarom is dit van kritieke belang om 'n geskikte metodologiese benadering vir impak-evaluering van maatskaplike programme in Suid-Afrika te ontwikkel. Met hierdie doelwitte, het die studie toepaslike metodologiese benaderings ondersoek ten einde beter program impak-evaluerings te ontwikkel deur die verkenning van die potensiële waarde van die Realistiese Evalueringsmetode (REM) op impak-evaluerings binne die Suid-Afrikaanse openbare sektor.

Die navorsing het 'n drieledige benadering gevolg. Eerstens, is 'n omvattende literatuuroorsig gedoen om 'n in-diepte begrip van die REM te kry deur middel van 'n gedetailleerde oorsig en ontleding van die verbandhoudende literatuur asook om die huidige tendense in navorsing en toepassing van REM benadering in die uitvoering van impak-evaluerings te beoordeel. Tweedens is ‘n mikro-ontleding van die gevallestudies voltooi om die metodes en benaderings te ondersoek wat in die verlede gebruik is vir program impak-evaluerings. Die gebruikswaarde van die evalueringsresultate is ook bepaal ten opsigte van nuwe insigte oor wat werk, vir wie, onder watter omstandighede en in watter opsigte. Laastens is in-diepte onderhoude met beleidmakers, kommissarisse en implementeerders van evaluerings gevoer om verdere inligting te bekom oor die gebruikswaarde van evalueringsresultate, sowel as die toepaslikheid van die REM as 'n metodologiese benadering in die uitvoer van program impak-evaluerings in die Suid-Afrikaanse openbare sektor.

Die algehele bevindinge het bewys gelewer dat aanvanklike goed-gedefinieerde en samehangende programteorie sowel as program oorsaaklikheid op impak-evaluerings grootliks afwesig is. Die navorsing het ook gelei tot beperkte kontekstuele begrip van die programme se interaksie met die breër komplekse sosiale sisteem. Hierdie bevindings bied beleidmakers beperkte insig in terme van begrip vir wie 'n sosiale program die effektiefste sal werk of nie, en die redes daarvoor. Daarom, het hierdie beperkte verklarende fokus gelei tot impak-evaluerings wat 'n "swart boks" verskynsel gehad het, aangesien die sleutel veranderingsmeganisme in die programme onbekend was.

Die gevolgtrekking is gemaak dat REM wel potensiële waarde kan bydra tot program impak-evaluerings ten opsigte van nuwe insigte oor wat werk, vir wie, onder watter omstandighede en in watter opsigte. Daar is egter potensiële beperkinge in die toepassing daarvan en dit moet goed oorweeg word teen die voordele wat verkry kan word uit sodanige evaluerings.

Een belangrike bydrae van die studie is 'n assesseringsmodel, van toepassing in die Suid-Afrikaanse konteks, wat toegepas kan word om vas te stel, uit 'n Realistiese Evaluering lens, wat die waarde van 'n evaluering sal wees vir die behoeftes van verskillende beleidmakers en toegepas kan word om die beperkinge te bepaal van ander impak-evalueringsmetodes.

Sleutelwoorde:

Impak-evaluering

Openbare sektor

Monitering en evaluering

Realistiese evalueringsmetode

Dedication

I dedicate this research to the memory of my late father, Duvalele Davis Mbava, and my mother, Eunice Nobabatizo Mbava. Your love, the solid foundation you laid in my life, the sacrifices and unwavering faith in my abilities have led to this moment.

To my three sons, Viwe Daniel, Zenande Gabriel and Lindile Raphael; this is for you. I have paved the way and the baton is now in your hands.

Acknowledgements

I wish to thank:

The Almighty God, for sustaining me and bringing me this far.

Lord, you establish peace for us. All that we have accomplished, you have done for us (Isaiah 26:12)

My life partner, Appleton Zuzile, for your love, encouragement and support throughout this journey.

My heartfelt thanks to my promoter, Dr Babette Rabie, for seeing value in this research, and choosing to walk this intellectual journey with me. Your passion, coaching, and commitment to excellence made this work possible.

The interview respondents, who opened their doors and hearts and engaged with me on this research. Thank you. Your insights and perspectives progressed the objectives of this research.

My editor, Mariette Nortjé, for her willingness to take on the massive editing task with great energy and commitment. Most of all, for caring about the quality of the final document.

Above all else, the National Research Foundation (NRF) is acknowledged for their financial assistance towards this research. Opinions expressed in this thesis and the conclusions arrived at, are those of the author, and are not necessarily to be attributed to the National Research Foundation.

Table of contents

List of tables

List of figures

List of appendices

List of key terms, acronyms and abbreviations

3ie	the International Initiative for Impact Evaluation
ANA	Annual National Assessments
CMO	context-mechanism-outcome (configuration)
CSG	Child Support Grant
DBE	National Department of Basic Education
DPME	Department of Planning, Monitoring and Evaluation
ECD	Early Childhood Development
EMIS	education management information system
GWMES	Government-Wide Monitoring and Evaluation System
M&E	monitoring and evaluation
MTSF	Medium-Term Strategic Framework
NDHS	National Department of Human Settlements
NDP	National Development Plan
NEP	National Evaluation Plan
NEPF	National Evaluation Policy Framework
NES	National Evaluation System
NRF	National Research Foundation
OECD-DAC	Organisation for Economic Co-operation and Development-Development Assistance Committee
PMG	Parliamentary Monitoring Group
RAMESES	Realist and Meta-narrative Evidence Synthesis: Evolving Standards
RCT	randomised controlled trial
RDP	Reconstruction and Development Programme
REM	Realist Evaluation Method
RSA	Republic of South Africa
SASSA	South African Social Security Agency
SNAP	Annual National Survey of Schools
StatsSA	Statistics South Africa
UISP	Upgrading of Informal Settlements Programme
UK	United Kingdom
UNICEF	United Nations Children's Fund
US(A)	United States (of America)

CHAPTER 1:
RATIONALE AND INTRODUCTION

1.1 INTRODUCTION AND BACKGROUND

Internationally there has been an evolution in the development of coherent country-driven monitoring and evaluation (M&E) systems and South Africa has followed suit. The drivers behind developing coherent institutional designed integrated national M&E systems have been the needs of transparency, accountability and the enhanced measurement of results and key policy decision-making in the public sector.

Cloete (2009: 308) observed that the establishment and implementation of national M&E systems have resulted in promotion of good governance and has impacted positively towards informing policy decisions. This resulted in the evolvement of country-led institutionalisation of M&E systems, which Gaarder and Briceño (2010: 4) described as "a process of channelling isolated and spontaneous programme evaluation efforts into more formal and systematic approaches." Effective and well-implemented institutionalised and integrated M&E systems inform and positively impact policy-making.

Segone (2008: 9-12) argued that policy-making should be informed by empirical knowledge emanating from monitoring and evaluation systems that supports strong evidence. The implication for evaluation is that systemic approaches towards impact evaluation maybe a requirement. Policy decision-making should be evidence informed and evidence-based. Such data should come from coherent and integrated monitoring and evaluation systems, which provide the evidence needed to take informed policy decisions.

In South Africa, institutionalisation of M&E systems is underpinned by the policy outlining the Government-Wide Monitoring and Evaluation System (GWMES), as ratified by government in 2007 (Republic of South Africa (RSA), 2007). According to Phillips, Goldman, Gasa, Akhalwaya and Leon (2014: 393), the advancement of an M&E system across government was anticipated to be a 'system of systems' where various government departments would contribute data and information emanating from their own systems to facilitate the availability of government intelligence. Essentially the government-wide M&E system data underpinnings were from secondary rather than primary sources. The system aimed to provide dynamic information of government performance to facilitate decision-making and policy review (Levin, 2009: 962). The overarching intentions of a GWMES data system was meant to integrate seamlessly the performance information emanating from government programmes, demographic and socio-economic statistical data as well as evaluation findings of various government programmes. Such rich and multi-faceted data was meant to provide credible and objective evidence that would aid in policy-making, planning and provide key information and insights supporting the budgetary cycle in the enhancement of National Treasury planning.

As a result of the disjointed nature of the GWMES, these aims were not immediately achieved. This view is supported by Cloete (2009: 297-298), who stated that poor coordination was a result of deficiency in policy and programme monitoring and evaluation. Phillips et al. (2014: 399) also indicated that the South African public sector had no apparent well-defined state planning mandate and embedded national plan. The lack of integration within the GWMES, as pointed out by Phillips et al. (2014: 393) and Cloete (2009: 297-298), resulted in central government departments generating their own single reporting systems with the resulting duplication of information requests and exacerbation of the reporting burdens.

The GWMES is underpinned by key transversal systems in the Presidency, Statistics South Africa and The National Treasury. These three 'data terrains' entail programme performance information; demography and socio-economic statistics as well as national evaluation. The GWMES policy framework is under the care and oversight of the Presidency. In practice, the National Treasury is leading the implementation of programme performance information and has issued the *Framework for Programme Performance Information* (RSA, 2007). Statistics South Africa (StatsSA) leads the implementation of social, economic and demographic statistics by publishing the *South African Statistics Quality Framework (2010)* (StatsSA, 2010). The third arm of the GWMES, the National Evaluation System (NES), came into effect in 2012 after the establishment of the then Department of Performance, Monitoring and Evaluation.

The national Department of Performance, Monitoring and Evaluation, established in 2010, later merging with the National Planning Commission to form the national Department of Planning, Monitoring and Evaluation, (DPME). The DPME proceeded to implement a National Evaluation System as guided by the National Evaluation Policy Framework, (NEPF), issued in 2011, which served to complete the three sets of policies that make up the GWMES (RSA, 2011a). This formalised the establishment of the National System of Evaluation to implement and provide oversight over public sector evaluations. With the full GWMES now firmly in place, it is expected that there will be full integration between National Treasury data planning demands and the reported data from the GWMES systems.

The *National Evaluation Policy Framework* identifies impact evaluations as one of the main types of evaluation the South African public sector will focus on. For a long time, the methodological approaches applied in impact evaluation have traditionally applied the randomised controlled trials approach prevalent in the medical field. This approach was quite successful in demonstrating whether a programme worked or not; however, it was found lacking in providing how and why a programme works or defining the 'programme mechanism'. Policy-makers were being short-changed and not fully getting the benefit of programme evaluation. As a result, policy-makers were left in the dark as to the identification of the key drivers of programme success, or lack thereof, and this has implications for programme replication in other settings. In order to judge a programme's impact, strong evidence of what works in programme efficacy has become a necessity.

This corresponds with the ongoing international debate in the field of evaluation on the appropriate methodological approaches for conducting impact evaluations. This debate centres around "the value and appropriateness of adopted methods and techniques and how the results and actual impacts of development interventions may be accurately attributed" (White 2009; 2010). White (2010: 155) argued, "There is a lack of evidence about what works and what does not – and at what cost.'' Other researchers identified research indicating that most evaluation findings have been based on limited or inadequate conclusions (Villanger & Jerve, 2009 in White, 2010: 156). In particular, demand has grown for evidence-based evaluations that demonstrate where, in what way, under what circumstances, and with what effects, programmes have worked (Greenhalgh, Wong, Jagosh, Greenhalgh, Manzano, Westhorp & Pawson, 2015; Lavis, 2009; Pawson et al., 2004).

Betts (2013: 250), on the other hand, argued that a range of methods available for synthesising evaluations has grown considerably and this profusion of approaches has offered little insight into which method might be most appropriate for marshalling a broad range of evidence on programme effectiveness and governance reforms with the aim of influencing policy reform. For this reason, theory-based evaluation approaches such as the Realist Evaluation Method have increasingly come to the fore as approaches that can potentially open the 'black box' of programme mechanism and provide greater insight on causality. It is contended that, since causal analysis in government programmes and policies is inadequate, the South African public sector requires robust and distinct programme planning and implementation phases that shed insights on what works (RSA, 2009: 21-22). The analysis of programme causality is currently insubstantial, and the international good practice of theory-based evaluation needs to be strengthened. The exploration of the efficacy of the Realist Evaluation Method is therefore opportune.

Pawson (2013: 15) asserted that "Realist Evaluation is avowedly theory-driven; it searches for and refines explanation of programme effectiveness." The overarching programme theory of change embedded in the Realist Evaluation methodological approach further describes the mechanism of how the programme causes change including describing the assumption, the risks and the ideal context that will result in those expected results. Weiss (1997c: 73) argued that, how change occurs is not the result of the activities of the intervention as such, but participants' reaction towards the activities of the intervention. Participants react to interventions in a rational manner and these rational reactions could be the intrinsic changes that result in the achievement of the envisaged outcomes. It has been found by some researchers that estimates of causal effects in random field experiments are nearly useless unless one also learns the mechanisms by which the causal effects are produced (Berk, 2011: 195). Therefore, Realist Evaluation opens up the 'black box' to gain better 'enlightenment' as to why and how the observed change occurred resulting in the net impact.

Debates on appropriate impact methodology often end in an impasse and deadlock between those advocating for impact evaluation methods that are primarily quantitative, with strong adherence to methodological purity and rigour (such as randomised controlled trials) and those advocating for methodological approaches that are qualitative with a view of providing a richer evidence base of what it is about programmes that works. Thus, this study contributes to the current thinking on how best to conduct impact evaluations by finding a middle ground of overcoming the quantitative versus qualitative nexus. Furthermore, as it supports robust impact evaluation that results in valid and adequate evaluation conclusions, this study can contribute towards reducing the stalemate between these paradigms.

Programme impact evaluation in South Africa should coherently and realistically indicate why, for whom, in what circumstance should a particular social programme achieve its outcomes. Therefore, a defined and appropriate menu of evaluation methods that provides choice to suit the evaluation objective at hand is necessary. According to Stufflebeam (2001: 9-10), such empirically tested methodical choices should be attractive in offering evaluators various methods which can be considered for appropriate application. This can serve to strengthen programme evaluation results that are far more conclusive. Such results can be supported by the utilisation of appropriate methods.

Notwithstanding, that the national Department of Planning, Monitoring and Evaluation (DPME) currently is in the process of embedding regular planned programme evaluations at national level, such expertise is largely external to government structures. The evaluation methods utilised are largely informed by the skills and experience of the commissioned practitioners and may lack the necessary rigour and systematic inquiry to adequately inform policy-making in an objective and evidence-based manner and thereby promote utilisation. As Mark, Henry and Julnes (1999: 179) argued, “the evaluator’s background may be the most important determinant of the type of evaluation that is done, rather than the context and the information needs of the affected groups and the public.” A further potential value of this study is therefore to highlight possible competency shortcomings in the producers of evaluation studies in South Africa.

The shortage of available impact evaluation expertise may in part account for the paucity of completed impact evaluations within the National System of Evaluation. The pipeline of planned evaluation, known as the National Evaluation Plan (NEP), feeding into the NES, currently has 54 evaluations that are completed and in progress (RSA, 2016a: ix). Appendix A includes all the evaluations that are in the NEP from 2011 to 2015. Current approved evaluations are shown in Appendix B and proposed evaluations are shown in Appendix C.

A review of the status of evaluations as at September 2015, reflected in Appendix A, illustrated that in the 2012/13 fiscal year, a total of eight evaluations were in the NEP. Two of these were impact evaluations, one was successfully completed and the other one was stopped. The following year 2013/14, a total of 16 evaluations were in the NEP. Five of these were impact evaluations, most of which were at draft report stage and were planned to be tabled at Cabinet for approval. In the

2014/15 fiscal year, a total of 15 evaluations were in the NEP. Five of these are termed 'impact/implementation' evaluations. Progress on these was varied with some at draft report stage, one evaluation stopped, other two evaluations at service provider selection stage and one evaluation approved by Cabinet. Finally, in the 2015/16 fiscal year, 11 evaluations are planned for in the NEP. Out of this total, only one evaluation is termed an 'impact/implementation' evaluation. This particular evaluation was also not implemented due to 'insufficient budget.

A summary of approved evaluation for the 2016/17 fiscal year as reflected in Appendix B, indicates that, nine evaluations are planned for, of which only one is an impact evaluation. Appendix C reflects proposed evaluation for the 2017/18 fiscal year. In this summary of proposed evaluations, there is no evidence of proposed impact evaluations as yet. Impact evaluations are clearly the second most under-represented type of evaluation research within the NES, with evaluation synthesis being most under-represented.

Therefore, a challenge is presented: whilst the National Evaluation Policy Framework identifies impact evaluations as critical, few are conducted and completed. It is argued that, since impact evaluations demand a greater level of skills and expertise, intense resource allocation, and are far more comprehensive and in-depth in nature, they are not easy to conduct. Whilst some impact evaluations are planned for in the annual evaluations plans, few are actually implemented and concluded. This possibly indicates the capacity and capability challenges of conducting these types of evaluations, as impact evaluations are arguably the most theoretically rigorous and resource intensive type of evaluations.

In addition to the above, evidence-based demonstration of programme impact is absent in those impact evaluations that are completed, as this study would argue in its conclusions. There is a paucity of evidence of what works and even less evidence on the reasons for success or failure. Internationally, policy-makers have been dissatisfied that research evidence is often misaligned with the policy-making cycle, that evidence is not strong enough and evaluation findings are inconclusive about what programmes work and how they work. This lack of empirical evidence of coherent programme impact leaves policy-makers in the dark on the possible success and reasons for such success on the adopted policies and programmes thereby constraining their ability to move to evidence-informed decision-making. Consequently, this international paradigm shift towards evidence informed policy-making, has resulted in international reforms towards more rigorous evaluations that support strong evidence of what works, for whom and under what conditions.

Addressing current capacity constraints with the evaluation of government programmes cannot be addressed by actors like the DPME alone. Locally, there has been M&E developments outside the government sector. These developments have seen a huge increase in M&E activities, training courses, formal studies and publications over the last decade as well as stable membership of the South African Monitoring and Evaluation Association (SAMEA) (Abrahams, 2015: 6).

SAMEA'S membership is largely made up of community of practice members from government, the private sector, civil society, non-profit organisations and higher education. The objectives of SAMEA are to develop M&E practices in the country through intellectual and thought leadership as well as promotion of M&E approaches and methods suitable for the South African context. A key and well-anticipated offering of SAMEA are the biennial conferences that attract a large turnout as emerging practices and current trends in the South African M&E landscape are pondered upon in this forum. These conferences and other initiatives receive support and collaboration from various partners including the Public Service Commission (PSC) and the DPME. SAMEA has standing memoranda of understandings (MOUs) with both the PSC and the DPME. Both these state institutions are legally mandated to monitor the performance of the public service and promote monitoring and evaluation practices for the benefit of the public sector. The respective MOUs were drafted in the spirit of cooperation, collaboration on M&E activities and practice in the public interest. Mutually beneficial initiatives include evaluation research, collaborations, M&E capacity building, joint hosting of workshops, seminars and conferences as well as education and training.

Other prominent players in the M&E landscape have been activist and interest group organisations largely outside government. Mouton (2010: 57) and Wildschut (2014: 254) have extensively studied the evolution of non-governmental organisations (NGOs) in the South African environment. Additionally, research by Abrahams (2015: 2) found that the nature of NGOs has evolved in post-apartheid South Africa, from being adversarial in relation to government towards a more collaborative co-existence. In this regard, whilst some NGOs maintain their non-governmental stance others have morphed into non-profit organisations (NPOs) as they are partly funded or worked with government.

It was found that in the new dispensation after the end of Apartheid, whilst international donors fund some NGO's, a number are largely funded by South African corporates. Both these funders require programme evaluation as a condition of funding (Wildschut, 2014: 307; Mouton, 2010: 88). This has led to increasing focus on M&E in the NGO sector. While the actions of the non-government sector are critical, the progress is often slower than what is demanded by service delivery targets. The South African state has seen increasing mass mobilisation against poor service delivery and widespread public dissatisfaction with programme performance and public management. This strong public demand for fiscal accountability as well as transparency resulted in appeals for enhanced monitoring and evaluation of programme performance. In addition, enduring inconclusive results of public programme performance have compelled the state to be accountable through demonstrated results and impact. Such results are established through coherent programme evaluation that provides accountability on programmes' effectiveness and outcomes. Alkin and Christie (2004: 12) reiterated this fact: "The need and desire for accountability presents a need for evaluation. The importance of accounting for actions or for resources used in the conduct of programmes is particularly evident for programmes supported by government entities." In this context, the evaluation

of government programmes, implemented and supported with public money, is required in order to satisfy the requirements of public accountability and transparency as an integral part of good governance. Therefore, in South Africa, there is a strong focus on outcome accountability through the strengthening of national planning systems, citizen oversight on public performance data and robust engagements with service delivery partners (RSA, 2009: 13).

Against this backdrop, in the last few years there has been an increasing call for 'Made in Africa', and 'Africa Rooted' evaluation. There is a sensitivity that evaluation in Africa is still emergent, and methods and practices are largely drawn from the countries in the Global North. It is well known that evaluations in the Global South are largely informed by the adoption of methodologies emanating from North American and Western European traditions. To a large extent, international development aid agencies facilitated the entry of evaluation methods and practices into the continent. Therefore the argument advocated by African evaluators such as the panel which convened in Bellagio, Italy to deliberate the issue (Bellagio Report, 2013) is that there is a need for a perspective of "evaluation theory and practice that is grounded in African philosophical assumptions about the nature of reality, drawing from African perceptions of the nature of being ("I am because we are"), from African worldviews and belief systems and ways of knowing, and informed by Africans' evolving values and aspirations"(Bellagio Report,2013:12). Chilisa and Malunga (2012:32) further emphasise that "African-driven evaluation theory and practice can draw from the evolving post-colonial indigenous paradigm to articulate epistemologies and values of an African-driven evaluation". Whilst it is not clear whether this demand the definition of entirely new paradigms that are indigenously African or an adaptation, refinement and customisation of the predominant ones to suit the African context, there are indications that the latter provides a quick win towards these aims and the former requires deliberate and purposeful paradigm shift. Proponents of the adaptation and refinement of the Global North paradigms claim that "Africa-rooted evaluation paradigm would not contain substantive differences from the prevailing Western evaluation paradigm, but its purpose, focus, design and implementation would probably just be more sensitive to African cultural contexts and practices in order to achieve the most accurate and valid results"(Cloete, 2016:67).Therefore it is incumbent upon African evaluators – informed by practice and their own contextual conditions – to consider how existing methods and practices can be adapted to uniquely African contexts.

On the other hand, the indigenous paradigm has at its core the ontological, knowledge and value systems that emanate from the cultures, histories, philosophies of those marginalised by colonialism (Chilisa, Major & Khudu-Petersen,2017:327). Evaluation paradigms in this context, transcend adaptation or adoption of prevailing methods. Rather they encompass evaluations whose methods, design, processes and systems and implementation are indigenous to Africa and its people in all respects. Carden and Alkin (2012:111) considers the African Peer Review Mechanism, an approach of The New Partnership for African Development (NEPAD) as a significant example of an indigenous paradigm "because of the significant engagement of African agents and agencies in its design,

implementation, critique, and evolution". Building on such foundations and other knowledge systems demands enhanced intellectual leadership in defining appropriate methodologies and approaches that enhance and push the boundaries of knowledge, and contribute towards enhanced theory building. Africans have been compelled to make this critical contribution and frame evaluation approaches contributing a variety of African-driven theories that add new voices and perspectives (Carden and Alkin (2012:106). Whilst the African-rooted evaluation discourse is firmly on the agenda and progressing, there is cognisance that its sustainability will demand a deliberate resolve as "we are still facing an uphill task in translating these efforts into widespread practice, especially on the continent, as the evaluation knowledge and practice gatekeepers are still mostly from the North (Chilisa & Malunga 2012:33).

Given these contextual factors, the monitoring and evaluation landscape in South Africa is dynamic and in continuous refinements. The evolution of programme monitoring globally and the initiation and establishment of programme monitoring in the South African public sector with the concomitant embedment of a national system of evaluation have been prominent. Notwithstanding this, there is increasing mass mobilisation and public dissatisfaction with programme performance and public management. Therefore, in the current context and within these identified problems and most critically the enduring inconclusive results of programme performance, the role of programme evaluation and appropriate evaluation methods offering new insights is important. The continuous search for better ways of judging the merit, success and impact of an intervention and the subsequent influence of such findings on policy- and decision-making in public management are critical and hence the role and value of this research in supporting those aims.

1.2 RESEARCH PROBLEM

The historical basis of evaluation and its evolution in the South African sphere has been well documented by Mouton (2010). Evaluation came into prominence in the aftermath of the Second World War when the US federal government embarked on social reform change initiatives such as the 'Great Society' and the 'War on Poverty' that included public health, social welfare and education programmes. This resulted in large budgetary spending on social programmes, spending which had to be systematically reviewed and justified. This then resulted in the growth of programme evaluation Mouton (2010: 10-12).

In this context, the quantitative evaluation methods that were prominent in the medical field such as randomised control trials were also applied as evaluation methods in social programmes. These quantitative methods of programme evaluation were influenced by the work of Campbell (1957) and Campbell and Stanley (1966) who defined the appropriate conditions for conducting experimental and quasi-experimental designs. Randomised controlled trials entail the measuring of the effect of a treatment between two groups. One group is a control group who receives the treatment and another group is a comparison group or a counterfactual, who does not receive the treatment or is given a

placebo. The observed results are then measured after the treatment to appraise efficacy of the administered treatment. Inferences are then made on the effectiveness of the treatment in comparison to the control group. According to Bamberger, Vijayendra and Woolcock (2010: 1) this then resulted in “efforts to adopt the standards and methods of bio-medical clinical trials in making knowledge claims about the effectiveness of particular interventions”. Therefore, impact evaluations have traditionally employed randomised controlled trial (RCT) methods as applied in the health sciences and the findings from such impact evaluations have largely been inconclusive because evaluation results of experimental methods are most effective in answering the question of whether the treatment worked or not. Whilst this method was well suited to laboratory experiments its application in measuring the success of social interventions was increasingly found unsuitable (Stufflebeam, 2001: 26; Pawson, 1997a; Pawson, 2013: 19 Woolcock, 2013: 1; Davis, 2000; Auriacombe, 2013: 719).

Woolcock (2013: 1) suggested that “the heightened focus on RCTs as the privileged basis on which to impute causal claims in development research and project evaluation has been subjected to increasingly trenchant critique in recent years”. This was due to limitations such as the validity of randomised controlled trials in programme evaluations, the inconclusive nature of experimental design methodologies in answering the question of why an intervention worked in a particular manner as well as difficulties in maintaining the methodological purity conditions such as the internal validity of a social intervention as well as the generalisability of the results to the wider social context. These and other methodological challenges of applying experimental methods in programme evaluations have been amply described by Davis (2000: 262). The literature further indicated that RCT evaluations do remain relevant in answering what happened and whether the intervention did work and should be selected in suitable evaluation circumstances; however, they are not practical or desirable in all situations

Under these circumstances, there were calls for programme evaluations that clarified how change came about by explaining what happened, how it happed and why. Stame (2004: 58) contends that evaluations have been overwhelmed by the ‘black box problem’ where only input and outputs are measured without due consideration of how and why change happened and this is “hardly informative for a policy design wishing to build upon previous experience”. Random control trials and quasi-experimental design methods largely tell us what happened after a treatment had been administered. They largely fail to answer the key question of why the treatment worked in a particular manner. The literature terms this the ‘black box phenomenon’. Questions regarding what it was about the experiment that caused the observed results or why the expected results were not observed remained unanswered, leading to the black box phenomenon.

This resulted in the development of theory-based evaluations. These were evaluations that specified the programmes theory of change that serves to demonstrate explicitly and logically the change mechanism of an intervention. The pioneers of this evaluation method include Chen (1990; 1994),

Rossi and Freeman (1985; 1993), and Weiss (1990; 1995; 1997a; 1997b; 1997c; 2000; 2004a; 2004b; 2008) who have collectively established thought leadership on the specification of the theory of change embedded in programmes and advocacy for evaluations testing such theories.

In this context, there is cognisance that social programmes are implemented in a complex reality. Stame (2004:63-64) argues that, the context under which programmes are implemented is stratified and multi-faceted with actors and agents embedded in their context. Unlike in 'simple' programmes where the input, outputs and resulting outcomes are linear, following a single causal strand, complex programmes have multiple strands which makes it difficult to evaluate all multiple strands as well as identify a single causality. This complexity is reinforced, according to Pawson (2006: 31) by contextual factors such as the individual capacities of the key agents and actors who have to enable the implementation of the social programme, the relationships between all programme key stakeholders, the institutional setting of the implementing agency as defined by its organisational culture as well as the overarching infrastructural system that supports the programme such as political backing, resource allocation, positive public perception and support. In this context, research has found that theory-based evaluations are promising in providing enlightenment in 'complex' programmes as theory-based evaluations can analyse causal mechanisms (Stern, Stame, Mayne, Forss, Davies & Befani, 2012: i).

Against this background, the *South African National Evaluation Policy Framework* (NEPF) is implemented towards the institutionalisation of a national system of evaluation. Its overarching aims are to define evaluation standards, methods and practice in the South African public sector. It aims to foster and support excellence in the management of evaluations, which are utilised for acquiring best practice in order to advance the efficiency and impact of public programmes, by interrogating what programmes are functional and effective and continuously adjusting interventions to achieve maximum effectiveness. It also seeks to confirm that sound and independent evidence from the evaluation results is utilised across planning, budgeting, organisational improvement, policy review, as well as programme implementation to improve performance (RSA, 2011a: iii).

The NEPF (RSA, 2011a: iii) defines evaluation as:

> *The systematic collection and objective analysis of evidence on public policies, programmes, projects, functions and organisations to assess issues such as relevance, performance (effectiveness and efficiency), and value for money, impact and sustainability and recommend ways forward.*

The framework further advocates that a "range of methodologies may be appropriate, and … a wide range of research techniques and data sources can be used, depending on the evaluation object and the evaluation questions at hand" (RSA, 2011a :6). Therefore, the National Evaluation Policy Framework is relatively receptive to all techniques and methods employed in the evaluation of programmes. The National Evaluation Policy Framework prescribes six evaluation types to be supported by the South African government, namely: (i) diagnostic evaluations; (ii) design

evaluations; (iii) implementation evaluations; (iv) impact evaluations; (v) economic evaluations; and (vi) evaluation synthesis.

However, according to Porter and Goldman (2013: 8), while impact evaluation of programmes is desired, there is a skills shortage in the South African public sector and "consequently, innovative methodologies are needed, the skills for which may be lacking". This has led to few impact evaluations executed, and those that have been implemented have been based on very simple questions such as 'what happened' without interrogating, 'how and why it happened'. This view is supported by Cloete (2009: 297) who claimed that evaluations have not always been undertaken or have been undertaken in superficial ways because of the applied research nature of conducting evaluations.

Consequently, there have been calls by government (RSA, 2009: 21-22) for the review and strengthening of theory-based evaluation approaches that articulate a "clear, conceptual understanding of how, why and when the policy, programme or project will effect change, and how these changes may be measured". Internationally as well there are calls for varied methods such as the view of the United Kingdom (2009: 14) Department of International Development which succinctly articulated this rationale:

> *Some development practitioners and researchers have promoted impact evaluation through experimental methods and randomised control trials as carried out in medicine. We recognise the usefulness of this type of work and support an increase in rigorous impact evaluations more generally, but as one tool in the evaluation toolbox, which must sit alongside evidence gathered through other evaluation methodologies.*

Stufflebeam (2001: 9) echoed a similar view stating "the study of alternative evaluation approaches is important for professionalizing programme evaluation and for its scientific advancement and operation". The Realist Evaluation Method proposes a more integrated approach to impact evaluation in terms of articulating both what happened and why it happened, and therefore the approach possibly present a better and more useful 'recipe' for evaluation. Realist Evaluation offers causal analysis of programme efficacy by unearthing the evidence of 'what works, for whom, in what context and in what respects'. Pawson (2006:25) emphasised that in Realist Evaluation the three elements of context, mechanism and outcome must be considered in order to answer the question of 'what works'.

Therefore, it remains important for the South African public sector to articulate and demonstrate credible results and actual evidence of impacts of development interventions. This has implications for improvement in service delivery, developmental outcomes and better-informed policy decision-making that is evidence-based. It is equally important that evaluators have access to various evaluation methods as dictated by the unique evaluation demands at hand. As Mark et al. (1999: 178) emphasised, "a multiplicity of evaluation theories can provide flexible evaluators with a

rich menu of approaches from which one can be selected that fits well with the need and demands of a particular evaluation."

The problem that is identified by this study is that, with the field of evaluation theory still booming internationally, there is no prescribed guidelines as to which impact evaluation method, employed in a specific context and under which conditions, will render the most useful findings for policy-makers. Therefore, clarifying a suitable methodological approach for impact evaluation of social programmes is critical in South Africa given the current focus and emphasis on public sector evaluation by the DPME, the NEPF and other government and non-government actors. Towards these aims, this study explored appropriate methodological approaches in informing better programme impact evaluations through the exploration of the potential value of Realist Evaluation on impact evaluations within the South African public sector.

The observation from Pawson and Tilley (1997a: 147) is illustrative in highlighting the research problem:

> *Evaluation reports simply indicating whether or not there has been a change associated with the introduction of a programme should not be commissioned or accepted by policy-makers. They are of no value, since nothing can be learned from them about what and what not to do in the future. Evaluation reports must identify not only the changes associated with the introduction of a programme but also what brought them about.*

Additionally, Chen's view (1994: 234) provided an apt rationale for this study:

> *We have lots of evaluation theory, but relatively little empirical knowledge of under what conditions the theory is valid or applicable, or under what conditions the guidance provided by the theory is useful or not useful in solving actual evaluation problems.*

1.3 RESEARCH QUESTION AND OBJECTIVES

The research explored and evaluated whether the Realist Evaluation Method (REM) offer potential value in the South African public sector as an additional method in programme impact evaluation. The Realist Evaluation conceptual framework was adopted to assess existing commissioned South African government programme impact evaluations on their adopted methodologies, with the aim of answering the main research questions, namely:

What is the potential value of adopting a Realist Evaluation Method in programme impact evaluations in the South African public sector? Can such an approach offer new insights of what works, for whom, under what conditions and in what respects, thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making?

The specific objectives are:

i) To gain an in-depth understanding of the Realist Evaluation Method through a detailed review and analysis of the related literature.

ii) To assess the current trends in research and application of Realist Evaluation methodical approach in conducting impact evaluations.

iii) To investigate the methodologies and approaches used in past programme impact evaluations in the South African public sector.

iv) To establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.

v) To establish the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector.

1.4 MERIT OF THE RESEARCH AND PROPOSED CONTRIBUTION TO SCIENCE

The purpose of this research is to explore the potential value of the Realist Evaluation Method in impact evaluations in South Africa in terms of providing new insights and understanding of how and why change happens in programme evaluation so that the results of evaluation findings are meaningful, have utility value and aid in policy-making.

The study will contribute greater knowledge towards the strengthening of South Africa's National Evaluation System by exploring whether the Realist Evaluation Method and its theory underpinnings can bring new knowledge and further insights and understanding on impact evaluation of government programmes in the South African context. It is envisioned that the research findings can address the following identified gaps and thereby offer the following contributions to the evaluation field:

Firstly, researchers such as Wildschut (2014) and Abrahams (2003: 268) who have engaged with theory-based evaluation methods have found a need for further research of this approach in a South African context. This study will address the current gap by constructing an in-depth understanding of the Realist Evaluation Method through a detailed review and analysis of the related literature and provide theoretical understanding of the method. A theoretical contribution to knowledge is the consolidation of an assessment model, **The Realist Evaluation Impact Evaluation Assessment Model** that adopts a Realist Evaluation perspective, which can be applied to assess the limitations of other impact evaluation methods and assist in deciding when a Realist approach will be useful and valuable to adopt. As Realist Evaluation is sensitive to programme context, the model has been adapted to contextual conditions prevalent in the South African monitoring and evaluation environment. This is relevant to the discourse on Africa-rooted evaluation paradigm which requires sensitivity to African cultural contexts and practices in order to achieve the most accurate and valid results. This strengthens the importance and significance of this study.

Secondly, other researchers, such as Cameron, Mishra and Brown (2016: 19); Rogers and Peersman (2014: 86); Mouton (2010: 184) as well as Coryn, Hattie, Scriven and Hartmann (2007: 438) have pointed to a need for further research on appropriate programme evaluation in the public sector. This study assists with this gap by investigating the methodologies and approaches

used in past programme impact evaluations in the South African public sector, will explain both the successes and shortcomings of such methods with a view to strengthening future impact evaluation in the public sector and the National Evaluation System at large. Therefore, a key theoretical contribution to knowledge is a synthesis of impact evaluations practice in South Africa. Currently such extensive synthesis on current impact evaluation practice in South Africa has not been conducted and this contribution provides insights on the current environment and the impact on the National Evaluation System.

Thirdly, with the increasing emphasis on 'evidence informed policy-making' (see Head, 2016; Oliver, Lorenc & Innvaer, 2014; Segone, 2008) this study contributes by establishing the utility value of the evaluation results, uncovering the needs of policy-makers before and subsequent to the evaluation, as a counter-point to what was actually done by the evaluators. Therefore, one of the key defined contribution of this study is a synthesised set of criteria of what policy-makers deem to be the most critical aspect required in a policy evaluation. The presence and articulation of these key aspects in a policy evaluation qualify such an evaluation to pass the acid test of being meaningful, valid and useful. It can thus be said that, (Pawson & Tilley, 1997a: 16) "the mandate comes from the policy-maker and the sensitive, experienced researcher selects the appropriate tools from the available tool kit." Therefore, this research aimed to meet these expressed needs of the policy-makers by exploring a suitable methodological approach for programme impact evaluation as an additional menu item that is supported by evidence.

Finally, establishing the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector is a key contribution to knowledge for the National Evaluation System. Therefore, a strong theoretical contribution of this study is the potential value of the Realist Evaluation Method in addressing some of the gaps and limitations that are evident in government programme impact evaluation. Such knowledge will enrich the community of practice and decision-makers with additional useful methods in the methodological toolbox of evaluation. This will assist in better impact evaluation of the social programmes of government and strengthen evaluation in South Africa.

This research made an adequate scientific contribution to these existing knowledge gaps. It is envisaged that the research findings will provide an element of 'enlightenment' on the contextual factors that might be suitable for the choice of this methodology in impact evaluation. As Patton (2008:31) reflected, there are various models, techniques, methodologies, philosophies and practices within the field of evaluation. Therefore, the challenge remains one of how to conduct effective evaluations amongst such competing and contradictory views.

1.5 SCOPE OF STUDY

The study explored the applicability of the Realist Evaluation Method on impact evaluations of social programmes in South Africa. The study explored social programme impact evaluation within the

South African public sector and excluded the realm of non-governmental organisations, non-profit organisations, international donor organisations and multi-lateral organisations. Specifically, public sector-led programme impact evaluations commissioned by the South African government (as opposed to donor-driven programme impact evaluation) are within the scope of the study.

1.6 OVERVIEW OF CHAPTERS

Chapter 1 provides the introduction and context of this study. The evolution of country-driven M&E systems is discussed and South Africa's path in charting its course in institutionalising its own M&E system is reviewed. This gives rise to the NEPF, whose purpose and main types of evaluation it espouses, is pondered. Appropriate methodological approaches in impact evaluations that can support evidence-informed decision-making are extensively discussed. Emanating from this background, the research problem, question and objectives, the merit of the research as well as the scope and limitation of the study are considered. Chapter 1 also presents a concise overview of the various subsequent chapters of the study.

Emanating from the research context detailed in Chapter 1, Chapter 2 begins by investigating trends in programme impact evaluation including the analysis of current theoretical positions and practice in both the global and local South African context. The review further progresses by discussing the rationale for evidence-based policy-making and the wider evidence-based policy debate on the systematic review of evidence. The trend towards review and synthesis of evidence-based policy-making is explored and an argument is made how such reviews and synthesis applying the Realist Evaluation approach can increase understanding about programme mechanisms and add insights that add greater value to policy-makers. Emanating from that discussion, the nature of evaluation and its objectives are discussed based on the literature analysis.

This then gives rise to a discussion on the enduring paradigm of experimental design methodologies in the evaluation of social intervention, and an extensive discussion on the long-standing 'paradigm wars' and the utilisation of experimental methods as the 'gold standard' in conducting programme evaluations with a view to gain evidence of methodological gaps and limitations. The current trends in global programme impact evaluation as well as the current situation in the South African context are pondered with a view to find evidence of methodological shortcomings and gaps prevalent in current programme impact evaluation and emerging best practice. Within this discussion, the current M&E landscape in South Africa both inside and outside government is discussed with a view of completing a synthesis of impact evaluation in this environment.

Informed by a theoretical base of the study in Chapter 2, Chapter 3 proceeds by providing the current context of the methodological approaches in impact evaluations globally. Within this backdrop, the emergence of Realist Evaluation Method is presented, including the theoretical foundations of the approach, the key ideas of realist inquiry and its research application. Its current application in the international public sector is discussed and examples from international application are highlighted.

Emanating from this background and critically evaluating the conditions under which to apply the Realist Evaluation approach, the review concludes by interrogating the suitability and limitation of Realist Evaluation with detailed information on the various practitioner and scholarly views about the method. Proceeding from this extensive review and the theoretical underpinnings of Pawson and Tilley's Realist Evaluation Method of programme evaluation, a Realist Evaluation Impact Evaluation Assessment Framework is advanced as a theoretical assessment lens to investigate the robustness of impact evaluations that are implemented in the South African public sector in terms of whether such evaluations are meaningful, valid and useful to policy-makers.

Chapter 4 proceeds to outline the overarching structure of the research design and adopted methodology that provides details of the research strategy and adopted data collection methods including the data analysis, research ethics and limitations of the study. Emanating from this, Chapter 5 presents the micro-analyses of the selected case studies applying the Realist Evaluation Impact Assessment Framework. Chapter 6 presents the data from the key informant interviews in line with two prime objectives of establishing from policy decision-makers, commissioners and implementers of evaluations in the South African public sector the utility value of evaluation results in offering new insights of what works, for whom, under what conditions and in what respects as well as establishing from the same respondents, the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector. Chapter 7 presents the consolidated research findings from the literature review, the assessment of the case studies through the lens of the Realist Evaluation Impact Assessment Framework as well as the opinions of the policy decision-makers and commissioners of impact evaluations, enabling the answering of the research question. The final section, Chapter 8 provides an overview of the research and the conclusions drawn based on the research findings. The importance and relevance of the research findings are discussed, as well as the significance of the work in contributing to the field of knowledge. In this regard a new model is advanced, **The Realist Evaluation Impact Evaluation Assessment Model,** which is more applicable to the South African context, that facilitates and streamlines prospective impact evaluations. Recommendations for the field as informed by the research findings are made. Finally, the limitations of the study are discussed.

1.7 SUMMARY AND CONCLUSION TO CHAPTER 1

This chapter outlined the rationale and context of the study. It presented the research problem, the research question and the objectives of study. The merit of the research and contribution to knowledge was discussed. The research scope was also further outlined. A summarised overview of the various chapters of the study was presented. The next chapter explores trends in programme impact evaluation including the analysis of current theoretical positions and practice in both the global and local South African context. This served as a contextual background for the evolution and

the emergence of the Realist Evaluation approach and its theoretical underpinnings, which will be extensively discussed in Chapter 3.

CHAPTER 2:
INTERNATIONAL THEORETICAL AND APPLIED TRENDS IN PROGRAMME IMPACT EVALUATIONS

2.1 INTRODUCTION

The aim of this review is to investigate trends in programme impact evaluation including the analysis of current theoretical positions and practice in both the global and local South African context. This served as a contextual background for the evolution and the emergence of the Realist Evaluation approach and its theoretical underpinnings extensively discussed in the chapter.

The review progresses by discussing the rationale for evidence-based policy-making and the wider evidence-based policy debate on the systematic review of evidence. Emanating from that discussion, the nature of evaluation and its objectives are discussed based on the literature analysis.

This then gives rise to a discussion on the enduring paradigm of experimental design methodologies in the evaluation of social intervention with a view to gain evidence of methodological gaps and limitations. The current trends in global programme impact evaluation as well as the current situation in the South African context are pondered with a view to find evidence of methodological shortcoming and gaps prevalent in current programme impact evaluation and emerging best practice. Within this discussion, the current M&E landscape in South Africa both inside and outside government is discussed with a view of completing a synthesis of impact evaluation in this environment.

2.2 THE SEARCH FOR EVIDENCE OF WHAT WORKS

The search for empirical evidence of what works has engrossed both researchers and policy-makers for some time. Researchers have been frustrated that research evidence is not influencing policy strong enough if at all. Policy-makers have been frustrated that research evidence is often misaligned with the policy-making cycle, is produced late and is often weak and inconclusive about what programmes work and how they work. This section aims to demonstrate and integrate the link between evidence-based policy-making, how the REM supports strong evidence of what works, for whom, under what conditions, as well as the approach's strengths in enhancing the review and synthesis of past evidence to inform policy-making. An overview of the prominence of evidence-based policy-making, the premised role that Realist Evaluation can play in providing strong evidence of what works and how the search for evidence has evolved towards research synthesis are discussed.

2.2.1 Evidence-based policy-making

Evidence-based policy-making is founded on the notion that public policy decisions should be both informed and grounded on empirical evidence in order to have the greatest impact. There has been an increasing emphasis placed on the use of research evidence to inform policy-making and practices. This is fuelled by the view that public policy should be built on strong and robust evidence, replacing the largely prevailing culture of judgement-based policy-making and practice (Davies, Nutley & Smith, 2000: 1-2). In this context, professional judgement was regarded with scepticism by an increasingly distrusting public, which demanded evidence of what works and accountability for public resources.

According to Taylor (2013: 6), evidence-based policy-making rose in prominence in the medical field in the 1970s through the replication of the results from RCTs within homogeneous groups. From then on, the 1990s saw such methods spread towards social policy by approximating the biological change observed in medical field with behavioural change in social programmes. Evidence is premised on empirically-derived knowledge which can be simulated and adapted in various contexts (Taylor, 2013: 5). Therefore, evidence-based policy premises the use of all available best evidence from various sources and past knowledge to inform policy decisions.

According to Segone (2008: 7), the concept of 'evidence-based policy' has become popular as a way of producing objective evidence to inform the policy process and result in better decision-making in government.

The prominence of evidence-based policy can be attributed to a citizenry that is politically aware and educated, the wide availability of information and development in information technology, the growth of a networked research community, and the demand for accountable government (Davies et al., 2000: 2). Rabie and Cloete (2009: 79) further asserted that the advent of information and communication technologies led to the incorporation of quantitative data analysis in social science which was previously informed by qualitative research methods due to data limitation. Therefore, innovation in information technology in the latter part of the 20th century enabled better and improved data analysis capabilities leading to informed evaluation research findings.

White (2010: 155) argued that there is a dearth of research evidence on how programmes work. Weiss, Murphy-Graham, Petrosino and Gandhi (2008: 31) identified research that found weakness in evaluation research as a result of questionable evidence, inability to meet the policy requirements of decision-makers, disintegrated programme data, inadequate evidence and inconsistent findings. White (2010: 156) concurred with this view and identified research, which found that most conducted evaluations provided inadequate evidence and invalid conclusions were drawn regarding programme impact.

Brown (2014: 20) identified research which found that non-utilisation of research evidence can result in squandering of public financial resources and the detrimental social exclusion of the susceptible underclass from benefiting from government programmes that are critical.

For this reason, the timeous provision of evidence to policy-makers or withholding of that evidence can potentially positively or negatively impact the life course of identified beneficiaries. Despite these factors, Weiss et al. (2008: 30) lamented the fact that evaluation seems not to have a strong influence on the policy process and is selectively used to fit the existing policy agenda. Weiss et al. (2008: 31) further pointed out that it is generally accepted that research evidence from evaluation has limited traction on policy influence. Pawson (2002a: 159) also argued that the underlying causes for this situation were primarily because evaluation research is not usually aligned to the policy-making cycle and hence there is incongruence between the research timing and policy sequence. In this regard, Weiss (1979, in Hawkes, Zaheer, Tawil, O'Dwyer & Buse, 2012) argued that since policy-making is complex, new knowledge and ideas from research evidence could provide decision-makers with 'enlightenment' in order to inform and shape policy-making. Enlightenment is also regarded as 'conceptual use' of evidence by policy-makers as such evidence shape the policy-maker's cognitive abilities in thinking about current and future policy options (Owen, 1992 in Johnson, 1998: 94).

2.2.2 Evidence synthesis and systematic reviews

For these reasons, demand has grown for programme evaluation that is evidence-based, demonstrating contextual conditions of programme success (Pawson, 2004; Greenhalgh, Wong, Westhorp & Pawson, 2011; Lavis, 2009). The use and application of 'scientifically-proven' evidence to inform the intervention strategies of policy-makers have been found to be critical.

Therefore, Pawson, (2002a:157) pointed out that systematic reviews of past experience, knowledge and evidence have been sought as means of garnering strong evidence to inform policy decisions. The trend towards review and synthesis of evidence-based policy-making is driven by dissatisfaction with the evaluation outcomes of random control experimental methods, which have been found to be inadequate in providing conclusive evidence on programme impact. Pawson (2002a: 158) emphasised that a strong base of evidence that has been systematically collected can be used to support or challenge policy and programme claims of what it is that works.

In essence, this review and synthesis of past successes and failures of social interventions can potentially provide policy-makers with first-hand experience of what has worked and infuse those lessons and best practices into future implementation of social interventions addressing various societal problems.

Pawson (2006: 7) added that for evidence-based policy to be credible, it must be underpinned by the marshalled reviews of previous interventions. This is effectively done when systematic reviews pertaining to a specific policy issue are interrogated, evaluated and synthesised in detail with a view to assess their validity (Davies et al., 2000: 7).

Pawson (2002a:160) pointed out that:

> *…there is nothing entirely new in the world of policy-making and programme architecture… in the era of global social policy, international programmes and cross-continental evaluation societies, one can find few policy initiatives that have not been tried and tried again, and researched and researched again.*

Therefore, systematic review and analysis of evidence on what has been done in bygone eras can shed light on what can potentially work in a particular policy implementation area. Segone (2008: 30) on the other hand posited that for evidence to influence policy and practice, it must be impartial, exhibit scientific rigour and be methodologically processed; in this vein, the evidence can be trusted and utilised as research evidence. Weiss et al. (2008: 32) agreed and pointed out that such valid and methodologically processed evidential data is credible as it is supported by transparency and facts and can improve the effectiveness of programmes and their replication in various contexts.

This balanced analysis of all available evidence on a policy issue enhances evidence-based and informed policy and represents the highest order of evidence. This is in contrast to sole research studies and the odd evaluation report which do represent the next level of evidence; however, they do not constitute systematic evidence since these odd studies may have methodological shortcomings and the results thereof are informed by their unique contextual factors which may not be open to replication or consistent in varied contexts. On the other hand, the methods of conducting systematic review (such as meta-analysis and narrative reviews) have been criticised for lacking an explanation about the mechanism of how a programme works, oversimplification of the programme outcomes and overlooking of the programme context (Pawson, 2002a: 163). The methods applied employ the use of classification categories, coding of variables across all programmes under review, resulting in numerical net effects of programmes' effectiveness. These methods have been found wanting as they oversimplify past programme mechanisms, context and outcomes. Weiss et al. (2008: 32) provided a case for the defence of systematic review and meta-analysis and argued that the view that these methods of assembling evidence strip away the programme context and yield inconsistent information on outcomes could be due to the fact that the foundational studies did not provide sufficient data on causal factors that characterised the programmes studied at the time and hence the original evaluations did not report on contextual factors.

Despite such identified shortcomings, Weiss et al. (2008: 32) argued that systematic reviews and meta-analysis remain powerful tools, not only for synthesising the outcomes of programmes over various environments, but also for recognising the key success factors that lead to the achievement of improved outcomes. Pawson (2002a: 179) on the other hand remained critical and further

contended that systematic reviews of past research that applied the realist methodical approach could better explain what works. Through the interrogation of programme mechanism, context and outcome patterns, Realist Evaluation offers an approach that is aware of, and articulates contextual circumstances of the programme and enables the replication of such interpretations. Therefore, realist explanation makes the context of programme evidence and outcome explicit, rather than offering decontextualised programme evidence and oversimplified outcome patterns as evident in the methodological strategies employed in systematic reviews. Realist review increases understanding about programme mechanism, context, and outcome patterns; insights that add great value to policy-makers as they provide nuanced understanding of 'what works, for whom, in what context and in what respects'.

It has been argued that complexity is at the core of programme efficacy and insightful articulation, appreciation and improvement of such complexity demands a good foundation of robust evidence (Pawson & Tilley, 2004: 14). Conclusions reached through evaluation are progressively constructed into research syntheses to inform programme efficacy (Cook 1997: 31 in Chelimsky, 1997).

The evidence-based approach contends that social policy has been based on opinion, beliefs, ideology and this tended to lead to policy decisions that the reflect personal values and beliefs of policy-makers and their constituencies; hence a call for impartial scientific research-based evidence to inform policy-making. The presence of evidence does not spontaneously lead to better policies as policies are essentially about the allocation of values, rather than the application of evidence. In line with this view, Oliver et al. (2014: 6) have been critical of the views that there is disjuncture between policy-making and the available evidence; that policy is typically not centred on available evidence and that usage of additional research evidence by policy-makers would result in 'better' policies. They claim these assumptions are impractical as there are different interpretations of what constitutes policy and therefore policy-makers may utilise other information sources other than what some methodologists regard as policy.

What constitutes evidence also remains contentious and there are contrasting approaches to what is knowledge and evidence generation that informs evidence-based policy (Hudson, 2001; Oliver et al., 2014; Davies et al., 2000: 2). Evidence can include the marshalling of facts from various sources including the adjudication from those who are knowledgeable and proficient on the issue, as well as anecdotal evidence. Evidence can emanate from various sources, the bulk of which is through academic research (Talbot & Talbot, 2015: 187). The consensus is that research evidence encompasses marshalling findings of systematic reviews so as to intensify the bulk of knowledge on a specific policy issue (Davies et al., 2000:3).

From the ensuing analysis, it emerges that evidence-based policy is underpinned by strong research synthesis. Evidence on programme mechanisms, the context under which the programmes succeed and programme outcomes are synthesised to find out the optimum conditions of programme effectiveness. These are essentially the building blocks of the Realist Evaluation Method and

therefore the REM is supportive and in alignment with evidence-based policy-making. The REM contends that for this to work optimally, research results should influence the thinking of policy-makers to enable better formulation of policy (Pawson & Tilley, 2004: 19). Therefore, Realist Evaluation findings can potentially bridge the gap between research and policy-making since Realist Evaluation findings can be far more practical and sense making than findings that are hidden in research jargon. The literature illustrates that, despite the compelling rationale for evidence-based policy-making, there remains a disjuncture between the utilisation of systematic research and policy-making due to various reasons. These reasons include the highly complex and non-linear nature of the policy-making cycle as well as limited utilisation of research output by policy-makers which have had implications for evidence-informed policy-making and implementation.

Therefore, given that there are calls that policies should be informed and based on empirical evidence in order to have the greatest impact, Realist Evaluation with its explanatory focus on how programmes work has the potential to provide a strong evidence base. This evidence base is enhanced by the systematic synthesis of evidence that characterises the building blocks of the Realist Evaluation approach. These building blocks construct the context of a programme, detail how change occurs and map the outcome patterns in a more explicit manner. Consequently, the building blocks offer evidence that adds huge value to policy-makers as they provide nuanced understanding of 'what works, for whom, in what context and in what respects'.

2.3 THE NATURE OF EVALUATION

Evaluation assesses programmes, projects, policies and other interventions in order to determine their merit, worth, or value. Scriven (2003/2004: 7) stated that:

> *…the evaluation process identifies relevant values or standards that apply to what is being evaluated, performs empirical investigation using techniques from the social sciences, and then integrates conclusions with the standards into an overall evaluation or set of evaluations.*

This definition is consistent with others that see evaluation as prioritising the appraisal of merit and value of a programme or policy (Stufflebeam, 2001: 11; House, 1993: 1; Scriven, 1997; 1999; 2003; Mark et al., 1999: 188). Intrinsic in this definition is the value judgement inherent in an evaluation. Fournier (2005: 140-141) stressed that evaluation inquiry and conclusions have a value judgement that is both empirical and normative. Mark et al. (1999: 188) supported the view of inherent value judgement as to the merit and value of a programme and its effect on the generation of new knowledge regarding programme efficiency.

Patton (2008: 39) described programme evaluation as:

> *…the systematic collection of information about the activities, characteristics, and results of programmes to make judgements about the programme, improve or further develop programme effectiveness, inform decisions about future programming and/or increase understanding.*

Therefore, these aforementioned definitions underscore the utility value of evaluation. Evaluation is pragmatic and has the ultimate goal of assisting in decision-making.

Others (Weiss 2004a: 154; Rossi, Freeman & Lipsey, 1999) see evaluation as a branch of applied research that employs disciplined enquiry and method use. Rabie and Cloete (2009: 81) concurred with this view and emphasised that "evaluation is nothing more than an applied social research activity." Weiss (in Alkin, 1990: 83) argued that evaluation is essentially research evidence on policies to enable better decision-making regarding conceptualising of future programmes. Lincoln and Guba (1986:73-84) further asserted that evaluation is 'disciplined inquiry' utilising scientific rigour on planned programming. This view is further supported by others (Rossi & Freeman, 1993: 5) who stated that evaluation is the scientific application of empirical social research in order to gauge how programmes were theorised, planned and implemented.

Therefore, the former definitions of evaluation focus on the function and purpose of evaluation, the latter definition reflects on systematic method use with theoretical orientation. Collectively these definitions of evaluation highlight that, evaluation is systematic analysis, supported by research evidence, about the outcomes of programmes and interventions in order to cast value judgement on their quality or worth, and the results ultimately assist in informing policies and decision-making. Therefore, these different views or definitions of evaluation and the lack of consensus illustrate the diversity of approaches to evaluation practice and theory.

Evaluation came into prominence during the 'Great Society' and the 'War on Poverty' social reform change initiatives that included public health, social welfare and education programmes. These initiatives strived to address social problems during the Kennedy and Johnson administration eras in the United States (US). Suchman (1967: 14) posited that the focus was upon programme development, the initiation of new services to meet demand and public need. These programmes were implemented *en masse,* and there were demands for these programmes to demonstrate their effectiveness. Consequently, this gave rise to the interest in using evaluation research to demonstrate the effectiveness of these programmes.

Alkin (2004; 2013) conceptually modelled an evaluation based on an Evaluation Theory Tree as depicted in Figure 2.1 below, consisting of use, methods, and valuing branches. Each branch represents evaluation theorists' thematic emphasis in evaluation. The 'use' branch represents evaluation theorists whose underpinning theory and orientation focus on utilisation of evaluation.

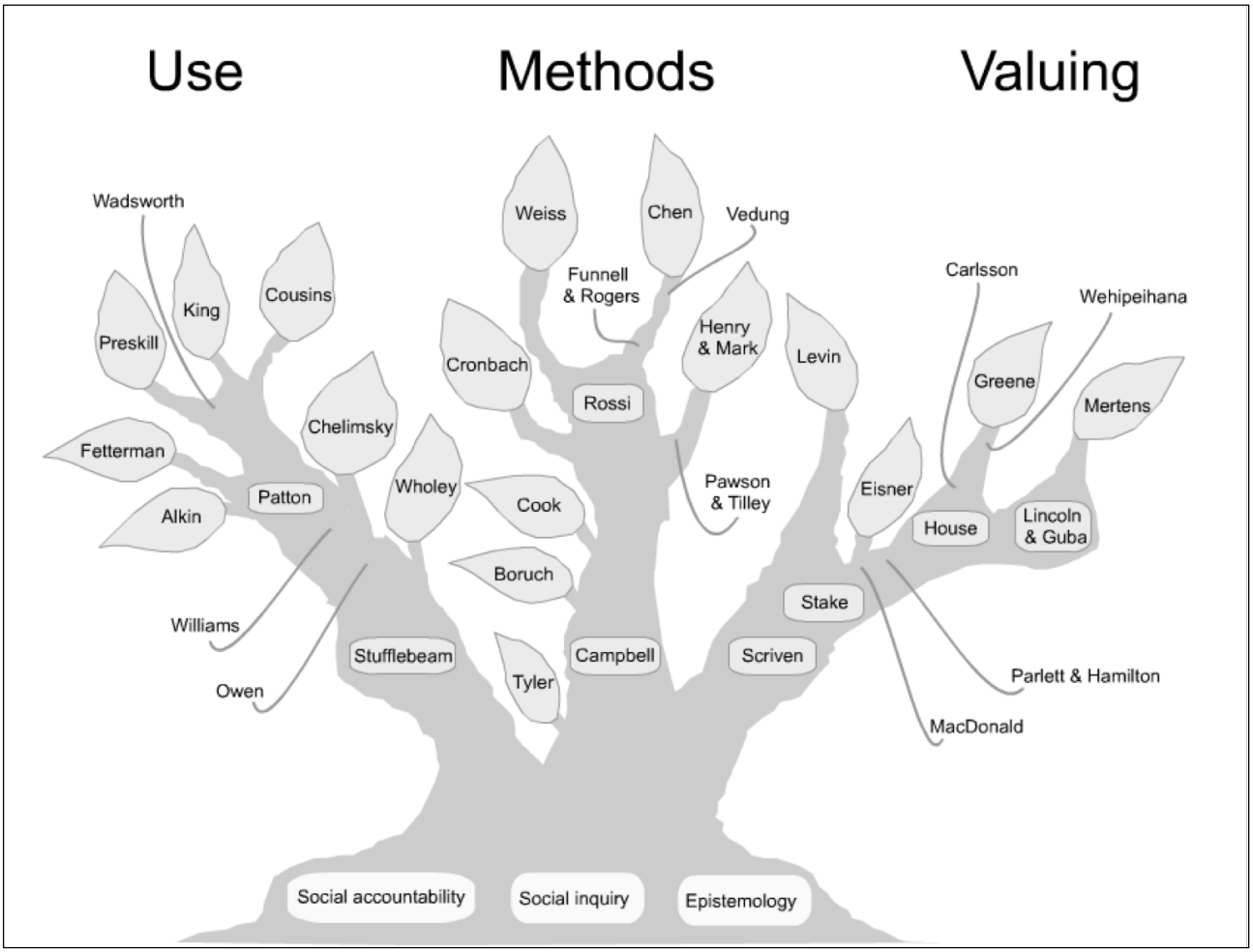


Figure 2.1: Evaluation Tree Theory

Source: Carden and Alkin, 2012: 105.

The foundational thinkers on utilisation are Patton (1984; 1986; 1990; 1996; 2008; 2012) and Stufflebeam (1966; Stufflebeam & Shinkfield, 1985; Stufflebeam, Madaus & Kellaghan, 2000). Subsequent theorists who have been influenced by utilisation-focused evaluation include Alkin (1990), Fetterman (1994) and Chelimsky (1997) and are represented by the leaves of the utilisation branch.

The method branch, as a theoretical orientation is influenced by the work of Campbell (1957) and Rossi and Freeman (1985; 1993). Weiss (1990; 1997a; 1997b; 1997c; 2000; 2004a; 2004b; 2008) and Chen (1989; 1990; 1994; 2005; 2010) went on to define theory-based evaluation built on this foundation. Accordingly, Realist Evaluation (Pawson & Tilley, 1997a), as a theory-based method is categorised under the theoretical branch. Pawson and Tilley have been designated as major 'European' theorists on the method branch of Alkin's (2013) modified conceptual 'evaluation theory tree' for significant influence and input to theory-based evaluation (Astbury, 2013: 383).

Carden and Alkin (2012: 105) further unpacked the value branch and its leaves. Valuing embodies the value judgement that these theorists espouse. Scriven (1999:521) emphasised that "programme evaluation is primarily concerned with judging the merit, worth, quality, or value of programmes".

This orientation towards casting a value judgement in evaluation is the overriding theme of theorists in the valuing branch. In addition to Scriven (1997: 477-501), Scheyer and Stake (1976: 39; Stake, 1991), House (1991; 1993; 2001), Lincoln and Guba (1986) also espoused this view.

Rabie and Cloete (2009: 78) supported this view and emphasised that evidence-based public policy analysis and innovations in social research methods have largely impacted the field of evaluation.

In conclusion, the overarching aspect of evaluation is that it embodies value judgement. There are varied theoretical methods applied in programme evaluation as influenced by the various thinkers. The evaluation tree framework conceptualised the various theories of evaluation where some thinkers subscribe to the utilisation focus of evaluation with the sole purpose of ensuring that evaluations are useful to stakeholders. Other thinkers in the 'valuing branch' believe evaluators have a role in casting value judgment regarding the worth or merit of a programme and this is intrinsically what distinguishes evaluation from research. On the other hand, those subscribing to the 'methods branch' have prioritised the precise application of scientific methods on social phenomena, seeking to preserve the internal validity of an experiment in order to eliminate bias. Theory-based evaluation arose from this foundation of the evolution of experimental method. Cardin and Alkin (2012: 104) support this view arguing that theorists who subscribe to the methods branch are influenced by the work of Campbell (1957) and Campbell and Stanley (1966) who defined the appropriate conditions for conducting experimental and quasi- experimental designs. Emanating from this background, the next section discusses experimental methods as conducted through randomised controlled trials and quasi-experimental methods.

2.4 EXPERIMENTAL METHODS

There has long been a paradigm divide and friction between those researchers preferring quantitative data research methods and those preferring qualitative data research methods. The nexus of this "great paradigm wars" is extensively discussed in Clarke and Dawson (1999: 37).

According to Pawson (2013:19), proponents of quantitative evidence held the view that qualitative data was gravely skewed and subjective, whilst the champions of qualitative evidence considered the application of quantitative data on social interventions as oversimplified. Woolcock (2013: 2) argued that RCTs, despite their limitations have been elevated to a 'gold standard' within impact evaluation, primarily influenced by powerful project designers, funders and those granting the project's initial approval in development agencies.

Experimental design methodologies, as reflected in Table 2.1, involve the measurement of the starting baseline before an intervention is implemented, as well as measuring the observed results post intervention. Administering the treatment entails a control group who receives the treatment and a comparison group or counterfactual, who does not receive the treatment. This process then provides an appraisal of the efficacy of the administered treatment with concomitant inference made on the effectiveness of the treatment in comparison to the control group.

Table 2.1: Experimental method

Random population	Pre-test baseline	Treatment administered	Post-test actual
Sample Group A: Experimental	Outcome A.1	Experimental group only	Outcome A.2
Sample Group B: Control	Outcome B.1		Outcome B.2

Source: Author.

Davies et al. (2000: 259) stated that judgement on the success of the experiment is limited to the sample only and the results of the experiment are cautiously generalised to the wider population.

Campbell (1957) pioneered the use of experimental methods and is acknowledged for his definitive groundwork on the elimination of bias on fieldwork research situations. His writing on experimental and quasi-experimental designs for research (Campbell, 1957) and his classic collaboration with Stanley (1966) studying experimental and quasi-experimental designs for research, have influenced the design of impact evaluations. The ideas put forth in Campbell and Stanley (1966) defined the foundation of social science research methods and contributed to how research designs were conducted for successive generations. They advanced new knowledge on how to effectively conduct RCTs within the social sciences and brought about the concept of internal validity for the internal control of an experiment and external validity in terms of assessing how the results were replicable in wider settings. Campbell and Stanley's work also illustrated that true experiments are not always ideal and that they cannot be used blindly. An alternative, such as quasi-experimental designs, was offered since quasi-experimental designs excluded randomisation as the treatment is administered to purposefully-selected comparison groups. Therefore, true experimental methods were quite effective in describing the effects and results of an administered treatment on defined outcomes and the condition and assumptions of true experiments were easily met in purely scientific environments; however, meeting such conditions and assumptions in field experiments of social programmes proved a challenge.

Auriacombe (2013: 719) observed that this hypothetico-deductive methodology of the scientific method based on an outsider perspective, stating a general hypothesis and proving or refuting the specific variables, means that evaluators applied statistical analysis to test social interventions. Suchman (1967) was influenced by Campbell's theories on experimental design and quasi-experimental design and pioneered the application of these theories in evaluation. In his definitive work, *Evaluative Research* (1967), he contributed to evaluation research and believed that evaluation should use rigorous scientific experimental design research methods.

In the 1970s, Cook, who had an interest in methodology, evaluation research, and social reform and contextual factors of evaluations, addressed issues and alternatives to random selection. He collaborated with Campbell, expanded on the foundational work of Campbell and Stanley by

advocating, and further developing the quasi-experiential designs where quasi-experiments involve pre- and post-test experiments and group comparison.

On the other hand, Boruch regarded random experiments as superior to any other statistical approach. In his writing (Boruch, McSweeney & Soderstrom, 1978; Boruch, 2005), he illustrated the random experiments and emphasised that results from random experiments have legitimacy as they are supported by strong scientific methods. His ideas on the management of evaluation were influenced by the work of Campbell's viewpoints that randomised field experiments were the most applicable and suitable approach in managing evaluations (Alkin & Christie, 2004: 24).

Boruch is a proponent of the use of randomised field experiments in the evaluation of social interventions (Boruch et al., 2000 in Alkin & Christie, 2004). He therefore claimed that social programmes should be evaluated scientifically, employing random controlled experimental methods to assemble authoritative and consistent evidence.

Other researchers such as Haynes, Service, Goldacre and Torgerson (2012:5) strongly promote the use of RCTs in public policy, claiming that these are cost effective and are "a powerful tool to help policy-makers and practitioners decide which of several policies is the most cost effective and also which interventions are not as effective as might have been supposed."

The 'development' of quantitative methods from the early 1960s in the sciences and their subsequent application in social research and social phenomena had implication for the field of evaluation as evaluation methods and approaches adopted this method. Quantitative methods, because of their supposed methodological purity, were deemed as a 'gold standard' in evaluation as they were thought to be superior to qualitative methods. Due to the criticism levelled against the application of quantitative methods on the social research, a countering paradigm of qualitative methods became prominent in the 1970s as other scholars and researchers advocated for qualitative research methods which were deemed relevant for the study of social phenomena. This resulted in a 'paradigm war' regarding the best methodology. This disagreement still endures and persists in influencing the viewpoints of some scholars and policy-makers.

Whilst methodologists such as Rossi were influenced by the experimental and quasi-experimental methods as pioneered by Campbell (Rossi, Freeman & Lipsey, 1999), they increasingly contributed to theory-driven evaluation of social programmes as underpinned by social research methods (Rossi & Freeman, 1985). This was the case with Weiss (1990; 1997a; 1997b; 1997c; 2000; 2004a; 2004b; 2008) who saw evaluation in a political context and as part of the policy research process, she advocated methodological sound research and thorough systematic inquiries adopting theory-driven approaches. Like Weiss, Cronbach's (1982) theoretical orientation in evaluation leaned towards policy research. He regarded evaluation as part of the policy research process. He therefore, advocated for evaluation use by policy-makers and saw evaluation as information that is provided to decision-makers so that they can make sound judgments of their own. He is regarded as one of the

great contributors on the methods evaluation branch including the 'Cronbach coefficient alpha' and 'generalisability theory' (Alkin & Christie, 2004: 30). In light of this state of affairs which still endures in the evaluation field, it is argued that the method that is adopted in an evaluation, either quantitative or qualitative or both should largely be driven by the objectives of the evaluation to facilitate the conclusive and valid answering of an evaluation question.

The US experienced the 'golden age' of evaluation of social programmes conducted in large part through the application of experimental methods from the sixties to the eighties (Davies et al., 2000: 259). During this time, there was an increasing review of the social programmes of the US 'great society' policies. These types of studies were quite prominent due to US government-driven demand for the evaluation of publicly-funded social programmes (Stufflebeam, 2001: 26). During the late 20th century there was increasing dissatisfaction with the evaluation results emanating from these methods. The traditional RCT experimental design and methodology, which have been prominent in programme evaluations, were challenged due to some identified limitations.

According to Davies et al. (2000: 254), RCTs are suitable for lab-based and inorganic study populations; however, applying this research approach to social interventions with human subjects is a challenge due to the complexity of social systems. Consequently, the method displayed limitations in answering the question of why an intervention worked in a particular manner.

Within the programme evaluation, despite the agreement that 'context matters' and that 'one size doesn't fit all' experimental methods have continued to trump the methodological standard. As Woolcock (2013: 1) contended, "champions of RCTs in particular imputed to themselves the moral and epistemological high ground as 'the white lab coat guys' of development research".

From the ensuing discussion, it is apparent that the evaluation results of experimental methods had limitations in terms of answering the question of whether the treatment worked or not. Questions about what was it about the experiment that caused the observed results or why the anticipated outcomes were not detected, remained unanswered, leading to the 'black box' phenomenon that has been strongly linked to these methods. Experimental methods were not practical to implement in all contexts due to difficulties and unethical reasons for developing the key counterfactual comparison group. Floyd, Pilling, Garner and Barrett (2004: 100) pointed out that there had been unswerving ethical misgivings and objections regarding the withholding of treatment from potential beneficiaries of a programme due to the fact that such beneficiaries happened to be randomly allocated to a control group who were not supposed to receive the intervention. However other researchers (Haynes et al, 2012: 16-17) argued that there are ethical advantages to utilising RCTs because some interventions that have been assumed to be beneficial have turned out to be harmful based on evidence from RCTs. Therefore, RCT in this context reduce harm and such experiments produce quality evidence that demonstrate the effectiveness of interventions.

Experimental methods have been criticised for their inability to probe the needs of the experimental group or offer a detailed understanding of the processes underlying the programme and therefore they exhibited limitations in providing substantial evidence regarding the merit and value of a programme (Stufflebeam, 2001: 26). Pawson and Tilley (1997a) contended, that since the evolution of the experimenting society as influenced by Campbell's 1963 seminal work on experimental methods in evaluation up until the late 20th century, evaluation research produced thus far had failed to offer high impact findings supported by strong evidence of what works. Pawson and Tilley (1997a: 10) argued that throughout these decades, the influence and power of evidence were highly contentious as it applied to whether a particular programme worked. As the experimental approach as practised through the decades was failing to offer answers on what works, proponents of this method simply called for more experimental research in the hope of capturing clearer programme outcomes.

The experimental approach's influence on the policy decision-making process was weak because the evaluation research produced was not useful to policy-makers and was therefore not utilised. The literature further indicates that experimental and quasi-experimental evaluations do remain relevant in answering what happened and whether the intervention did work and should be selected in suitable evaluation circumstances (see Nave, Miech, and Mosteller (2000) in Stufflebeam, 2001: 27). However, experimental and quasi-experimental evaluations are not practical or desirable in all situations. As a result, there have been calls for evaluations that have validity and utilisation value, incorporating qualitative elements. Pawson and Tilley (1997a: 22) surmised that the record of accomplishment of the experimental approach has been mixed under replication and have tended to "indicate programme success in some respects here but not there, and in other respects there but not here".

There have been debates focusing on methodological approaches to impact evaluation, the importance and suitability of applied evaluation methods and practices and whether proper causality and attribution were accurately determined (White, 2009; 2010). Yet, other debates have focused on whether to apply a qualitative or quantitative method. This qualitative-quantitative nexus has resulted in much mudslinging and the debates have become intense even fervid with strong adherence and critics of one or the other (Pawson, 2002a: 158; Chen, 1994: 231).

Patton (2008: 683 in Given, 2008) posited that both qualitative and quantitative approaches are progressively utilised in programme evaluation to open the 'black box'. Mixed-method approaches, employing both quantitative and qualitative methods in the same study, have been seen as a potential middle ground. However, Chen (1994: 232) was dismissive of applying mixed methods in an evaluation and likened such a strategy to a 'shotgun marriage' since quantitative and qualitative methods are diametrically opposed to one another and such a strategy lacks a conceptual framework. Later, Chen (2006: 75) conceded that such shotgun marriage can be averted and conflict between both quantitative and qualitative methods can be minimised if mixed methods are applied

under “the conceptual framework of programme theory”. The mixing of methods under this framework is undertaken with an overarching goal of strengthening and clarifying programme theory. Other researchers, such as Bamberger, Vijayendra and Woolcock (2010), argued the merits and demerits of using mixed methods research citing various challenges with the approach. Some of the challenges identified with the mixed methods research approach are that these evaluation designs lack some of the expected methodological and conceptual rigour and mixed method evaluations often results in two separate strands of quantitative and qualitative survey and data collection methods that lack cohesiveness (Bamberger, Vijayendra & Woolcock, 2010: 23-24). Therefore, whilst mixed methods present opportunities in the further strengthening of programme impact evaluation, their implementation presents some challenges.

On the other hand, whilst RCT remain the default evaluation approach in many parts of the world, they are not always feasible. Carden, Bamberger & Rugh (2009:1) argue that, the complex nature of many interventions impacted by resource constraints such as budgets, time and data, and the fact that the counterfactual comparison groups are not always possible in experimental design, requires alternatives to the established statistical counterfactual as an additional menu item in the evaluator’s toolbox. Ideas have been suggested including theory based approaches, quantitatively based approaches as well as qualitatively based approaches. Within theory-based approaches, programme ToC was found useful where a programme seeks to create change and incremental change can be tracked. Logic model were found useful where the programme had an explicit implementation strategy. Realist Evaluation was found useful in circumstances where the programme is implemented in different context with variations in outcomes.

In addition to these approaches other creative approaches included the application of historical, forensic and criminal investigation methods (Carden, Bamberger & Rugh, 2009:9). Quantitatively based approaches including pipeline design prevalent in large scale infrastructure projects; taking advantage of natural experiment which can provide an unanticipated but useful counterfactual as a result of delays in implementation; creative uses of comparison groups in the absence of a random selected group; comparison with other countries using similar measurable indicators as as well as commissioned surveys such as citizens reports and public expenditure tracking studies. Finally, qualitatively based approaches entails ideas such as concept mapping where this can be used to select a set of representative programmes or countries for preparing in-depth case studies. In addition, some of the quantitative techniques described can equally be customised to collect qualitative data (Carden, Bamberger and Rugh,2009:9).

2.5 PROGRAMME IMPACT EVALUATION GLOBALLY AND CURRENT TRENDS

There are varieties of programme evaluation and in this section, the specific discussion focuses on programme impact evaluation which refers to counterfactual-based programme evaluation that

attempts to attribute specific outcomes of interventions by dealing with the problem of selection bias (White, 2010: 157-158).

International development visually locates impact evaluation in the apex of the results chain as depicted in the 'logframe' or 'logical framework'. The logical framework analysis is an adopted methodology, which was put in place by many bilateral and multi-lateral development assistance organisations, and it serves to articulate the management of programme results (Carden & Alkin, 2012: 107). The 'road to results' (Morra-Imas & Rist, 2009 in Carden & Alkin, 2012: 107) further develops this approach in detail. Others, such as Wildschut (2014), have studied a systematic review tracing the development and evolution of the programme logical framework. This study focuses at the apex of the triangle where programme outcomes and impacts are interrogated. Figure 2.2 below depicts the logical framework analysis that provides a useful and visual way of mapping out programme management.

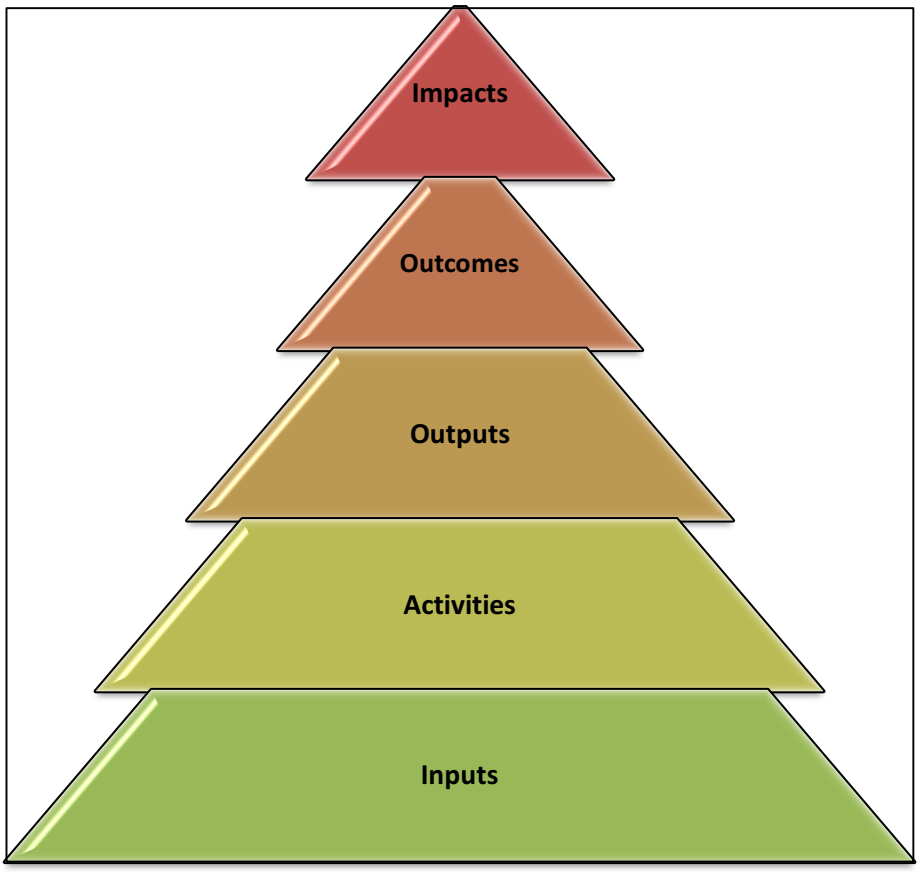


Figure 2.2: Programme Logic Framework

Source: RSA, 2011a.

Inputs (What is used to do the work?): the resources that are applied in order to achieve certain outputs, including personnel, finance, materials, time.

Activities (What is actually done?): the activities and processes that utilise certain inputs in order to achieve certain outputs.

Outputs (What is produced or delivered?): the products and services that result from the activities and the application of inputs.

Outcomes (What is intended to be achieved?): the medium-term results/benefits that are enjoyed by beneficiaries as a consequence of access to the delivered outputs.

Impacts (What is intended to be changed?): the meaningful developmental effect to which outputs contribute.

Research indicates that programme evaluation in its early history drew much of its theory and procedures from experiences in evaluating schools and school programmes (Stufflebeam, 2001:20). However, globally programme impact evaluation has been prominent in health nutrition and population, education, agriculture rural development and social protection and very little impact evaluation evidence exists of interventions in transportation, energy, economic policy and urban development (Cameron, Mishra & Brown, 2016). Accordingly, impact evaluations emphasise the generation of a broad evidence base in evaluating programme impacts and require strong evidence to make causal inferences. Secondly, this type of evaluation requires the interrogation of intended and unintended outcomes as guided by policy goals.

The Organisation for Economic Co-operation and Development-Development Assistance Committee (OECD-DAC) defines impacts as “positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended” (OECD-DAC, 2010: 24 in Rogers & Peersman, 2014: 86).

Rogers and Peersman (2014: 86) argued that impact evaluation seeks to make causal deductions about the causal nature of programme impact. Owen and Rogers (1999: 264) supported this view, adding that programme impact evaluation is concerned with assessing the footprint of programme outcomes, including the implementation nature of the programme, all of this conducted with the primary aims of delivering programme evidence to all relevant stakeholders and the decision influencing of future duplication, expansion or replication in other contexts. However, according to some, judging the value and worth of a programme or policy is equally important. According to Scriven (1999: 521), “programme evaluation is concerned to establish the merit, worth, quality, or value of programmes, in whole or in part, at the request of some client or clients, and for the benefit of some audience”.

Currently, concentrated effort has been on the ‘results agenda’ and ‘impact agenda’ driven by the Millennium Development Goals and the subsequent Sustainability Development Goals. The ‘impact agenda’ calls for programme impact evaluations that are relevant to policy-making and communicate evidence that provide clear policy direction. The key issue is to progress from the monitoring of programme outputs towards clear outcomes and impact. This has resulted in a clear focus on programme impact evaluation and is primarily driven by the increasing call for evidence of results on the performance of government programmes. This entails a major shift from monitoring input and

outputs, which are short term in nature towards a rigorous evaluation of long-term and lasting impact of programmes and other interventions. The pure measurements of outputs, as is normally the case in performance monitoring, offers little or limited insight about the realisation of government programme outcomes on a long-term and sustainable basis.

Given all these arguments, Rogers and Peersman (2014: 85) lamented that these disagreements in the field of impact evaluation are hardly informed by any form of strong and methodological research evidence.

A recent study (Cameron et al., 2016) mapped the growth of programme impact evaluation research conducted over time from the 1980s to the present and indicated that there was a precipitous rise in the number of impact evaluations conducted between the years 2000 and 2012, from 30 impact evaluation studies in the year 2000 to 370 studies 2012. Evidence points to this research emanating largely from health science journals followed by social science journals, international banks, research institutions, universities, NGOs, and a small percentage from governments in the form of research reports. Cameron et al. (2016: 8-10) further claimed that sectors conducting impact evaluation during this period were dominated by the health, nutrition and population sector (64.9%), followed by education (23.1%), social protection (15.1%) and agriculture and rural development (9.7%). In comparison, impact evaluations coming from the public sector during this period were 3.3 percent of the total. The evidence further indicated that randomised controlled trials as a method dominated the impact evaluation of health, nutrition and population at 83 percent, followed by water and sanitation services at 69 percent, whilst information and communications technology stands at 68 percent and education lags at 60 percent.

In-depth analysis indicated that most of the research evidence emanated from South Asia at 21.9 percent, a large proportion came from Eastern Africa at 19 percent, whilst South America produced 14.7 percent of research evidence and Central America including Mexico produced 10.7 percent. Impact evaluation evidence coming from Southern Africa is minuscule in comparison at 5.6 percent (Cameron et al., 2016: 12).

There has been an increased focus on impact evaluation in recent years within the public sector and the development sphere. Cameron et al. (2016: 18) highlighted that the public sector has been publishing far more frequently evaluation research evidence since 2009, specifically public sector-driven development interventions. Much of the discussion has focused on how to conduct better impact evaluations. It has been found that poor quality impact evaluations harm development by providing unsound, distorted and superficial findings which lead to poor decision-making (Rogers & Peersman, 2014: 86). However, the research indicated that government agencies produce impact evaluations at least once a year, from the time of research to final publication compared to the average time of six years in research journal publication and nearly four years in banks and international lending agencies (Cameron et al., 2016: 15). Therefore, the impact evaluation research commissioned from the public sector appears to be quicker and faster; however, the quality of that

output remains unverified and whether such findings are in fact conclusive and do influence or make an impact on policy-making are unknown.

On the other hand, within the international development sphere, global development finance institutions such as the World Bank, have defined and influenced the type of evaluation methods applied in much of the developing world. As a global multilateral finance institution, the World Bank, provides multi-billion finance loans to developing countries, to finance mega infrastructural and development projects. Evaluation within the World Bank has been critical in influencing its lending decisions. Much of the evaluation function within the bank is devoted to evaluating development operations in sectors such agriculture, energy, finance, transport and water supply. Evaluation is also used in instruments such as investment projects, financial intermediation schemes and structural adjustment policies. In addition, evaluation has been used to assess impact and process evaluations, country reviews and country programme assessments (Picciotto, 1997:203).

Since, the World Bank is the single largest provider of development finance in the developing world, its methods, ideas, and missions have largely been spread through large development research programmes, training provided to development policy makers, its development expertise, global professional networks and strategic alliances (Picciotto 1997:202). These ideas have had an indelible influence on the evaluation field.

The impact evaluation methods promoted by the World Bank are largely quantitative experimental design methods. The World Bank believes “randomised selection methods represent the strongest methods for evaluating the impact of a programme (Gertler, Martinez, Premand, Rawlings, Vermeersch, 2011:50). Therefore, the application of RCT methodologies in impact evaluations has been the hallmark of most World Bank evaluations. In addition, the World Bank has commissioned various research supporting the use of RCT in development effectiveness. Such findings advocated for the greater use of randomised evaluation methods in impact evaluation, arguing that “just as randomized trials for pharmaceuticals revolutionized medicine in the 20th Century, randomized evaluations have the potential to revolutionize social policy during the 21st century (Duflo &Kremer 2003:32).

Similarly, The Initiative for Impact Evaluation(3ie) founded in 2008, arose from the recommendations of a working group study of the Center for Global Development(CGD), a think tank, some of whose founders were affiliated with the World Bank. The evaluation working group identified an apparent ‘evaluation gap’ because “after decades in which development agencies have disbursed billions of dollars for social programs, and developing country governments and nongovernmental organizations (NGOs) have spent hundreds of billions more, we know relatively little about the net impact of most social programs” (CGD 2006:1). The role of 3ie in this regard was to close this gap and to assemble and systematically build a repository of evidence about what works in social development interventions globally, with the aims of providing evidence for policy-making, decision-making and learning across countries.

The rigorous impact evaluation methods that are promoted largely emulate the same methods that are applied in bio-medical fields as it is claimed that, “clinical trials of medications have become a standard and integral part of medical care. No physician would consider prescribing strong medications whose impact and potential side-effects have not been properly evaluated. Yet in social development programs, no such standard has been adopted” (CGD,2006:15).In this regard recommended methodical approaches to impact evaluation are quantitative experimental designs such as RCT as well as “controlled before-and-after studies, interrupted time-series studies, and various types of matched comparison studies (such as difference-of-differences and propensity score matching)” (CGD,2006:15). Such evaluation designs are deemed to provide rigorous evidence in decision making.

For the reason that 3ie consolidate, execute and provide thought leadership on what is deemed to be quality standards of rigorous impact evaluations, the organisation has an influence on the trajectory of impact evaluation designs particularly in the developing world. Most of these impact evaluations are implemented or supported by member countries, multilateral development institutions, research organisations, bilateral agencies, philanthropic foundations and nongovernmental organisations.

Therefore, both the World Bank and 3ie are powerful global organisations that influence the synthesising and dissemination of impact evaluation evidence from a particular and precise methodological stance. This implies the perpetuation, globally, of ideas and viewpoints that there are only particular methodological designs that characterise the most rigorous methods for evaluating the impact of a programmes. This further prolongs the methodological paradigms as well as the illusion of a gold standard in impact evaluations.

Therefore, programme impact evaluations are important in international development. Given this imperative, Betts (2013: 250) argued that a range of methods available for synthesising evaluations has grown considerably and this profusion of approaches has offered little insight into which method might be most appropriate for marshalling a broad range of evidence on programme effectiveness and governance reforms with the aim of influencing policy reform. Consequently, this research advances and explores whether Realist Evaluation might be a suitable method for providing strong evidence on what works in programme evaluation.

2.6 CURRENT PRACTICE IN SOUTH AFRICA

This section discusses monitoring and evaluation within the overarching context of the results-oriented public sector internationally as well as in the South African context. The centrality of programme impact evaluation within this context is highlighted supported by the emergence and current implementation of the National Evaluation Policy Framework.

2.7 SOUTH AFRICAN MONITORING AND EVALUATION ENVIRONMENT

International development organisations and governments around the world, have utilised monitoring and evaluation (M&E) to inform performance results and outcomes of development interventions. Many governments, including South Africa, are increasingly adopting M&E systems, processes and methodologies, traditionally practiced by donor agencies.

Within the overarching context of the results-oriented public sector, coherent institutional design of integrated national systems of M&E is a national imperative for governments globally. Towards this aim, best practice techniques, models and methods are utilised to assess public sector driven development interventions (Coryn, Hattie, Scriven & Hartmann, 2007: 438). Consequently, governments are also exploring emerging methods and approaches of assessing programme effectiveness.

Monitoring and evaluation practices in South Africa occur in the context of limited historical evidence of the effectiveness of public sector strategic programmes and those of significant public interest such as industrial developments, housing and rural development. Public sector programme evaluation has been absent either at the planning stage of interventions, throughout programme implementation, and subsequent to implementation. In light of this, the South African government is entrenching the process of institutionalising M&E planning frameworks. Some researchers (Mouton, 2010) have extensively documented the history and emergence of programme evaluation in South Africa, whilst others (Naidoo, 2011) have examined the link between public sector M&E and good governance in South Africa.

The current monitoring and evaluation environment across government in South Africa has largely been one of measuring activities, rather than the effective measurement of outcomes. There has been a concomitant dependence on process indicators rather than actual improvements and the final impact of change interventions. This has resulted in malicious compliance where set targets were not stretching enough and the performance reporting focused on processes, rather than results, which led to poor accountability and lack of evidence in the achievement of planned outcomes. Therefore, lack of quality M&E information was a further factor that constrained effective policy decisions pertaining to evaluation of government interventions, institutions, programmes and policies. Research by Rabie (2011: 451) found that the failure to enforce the use of performance information in planning and decision-making may give rise to M&E systems that are “toothless and result in no or minimal compliance, leading to another ‘good on paper’ policy that delivers little results in practice”.

There is now new emerging thinking regarding the effectiveness of the evaluation of public policies, programmes, projects, functions and organisations of state in terms of relevance, their performance in terms of efficacy, economy, and impact. The South African public sector is shifting its focus from the sole monitoring of programme outputs and is embedding systems and processes of monitoring

and measuring priority outcomes on a long and sustainable basis. According to Engela and Ajam (2010: 13), this was the underlying rationale for the establishment of an M&E oversight body such as the DPME in South Africa. The ministry was established to ensure that the results of massive government spending in basic services are clearly demonstrable through evident and sustainable outcomes. It has become essential to interrogate whether public sector programmes and plans are yielding the intended results, are indeed benefiting the intended beneficiaries, as well as producing the desired and expected outcomes. Hence, the importance of undertaking impact evaluations as a vital component of public sector accountability is emphasised.

A major development in the South African M&E field has been the establishment of a National Evaluation Policy Framework (NEPF) to inform a National Evaluation System (RSA, 2011a).

The scientific method of providing impartial, evidence-informed research information on programme efficacy is dominant within the NEPF. Intrinsically, political accountability is the driving force behind the emergence of evaluation in the South African public sector, leading to evaluation taking a prominent role in the NEPF rather than being circumscribed by performance management.

Until the implementation of the NEPF, evaluation as a discipline had not been fully institutionalised in the South African public sector and therefore, there is a vacuum and lack of tried-and-tested methods and practices on evaluation in the public sector. In this environment, flux with uncertainty and lack of proven experience, new procedures, practice standards, methodologies and competences should be developed and tested.

2.8 CURRENT CONTEXT OF PROGRAMME IMPACT EVALUATIONS

In South Africa, the monitoring and evaluation system is informed by an overarching Policy Framework of Government-wide Monitoring and Evaluation(GWMES) (RSA, 2007). The evolution of the GWMES and its implementation is extensively discussed elsewhere (Engela & Ajam, 2010). The GWMES, has three components that inform government performance monitoring: the management of social, economic and demographic statistical data across government, as well programme evaluation across government. These components are prescribed by the *Framework for Programme Performance Information* (RSA, 2007), *South African Statistics Quality Assessment Framework* and the *National Evaluation Policy Framework* (RSA, 2011a) respectively. The structure of this monitoring and evaluation system and its various components are depicted below in Figure 2.3.

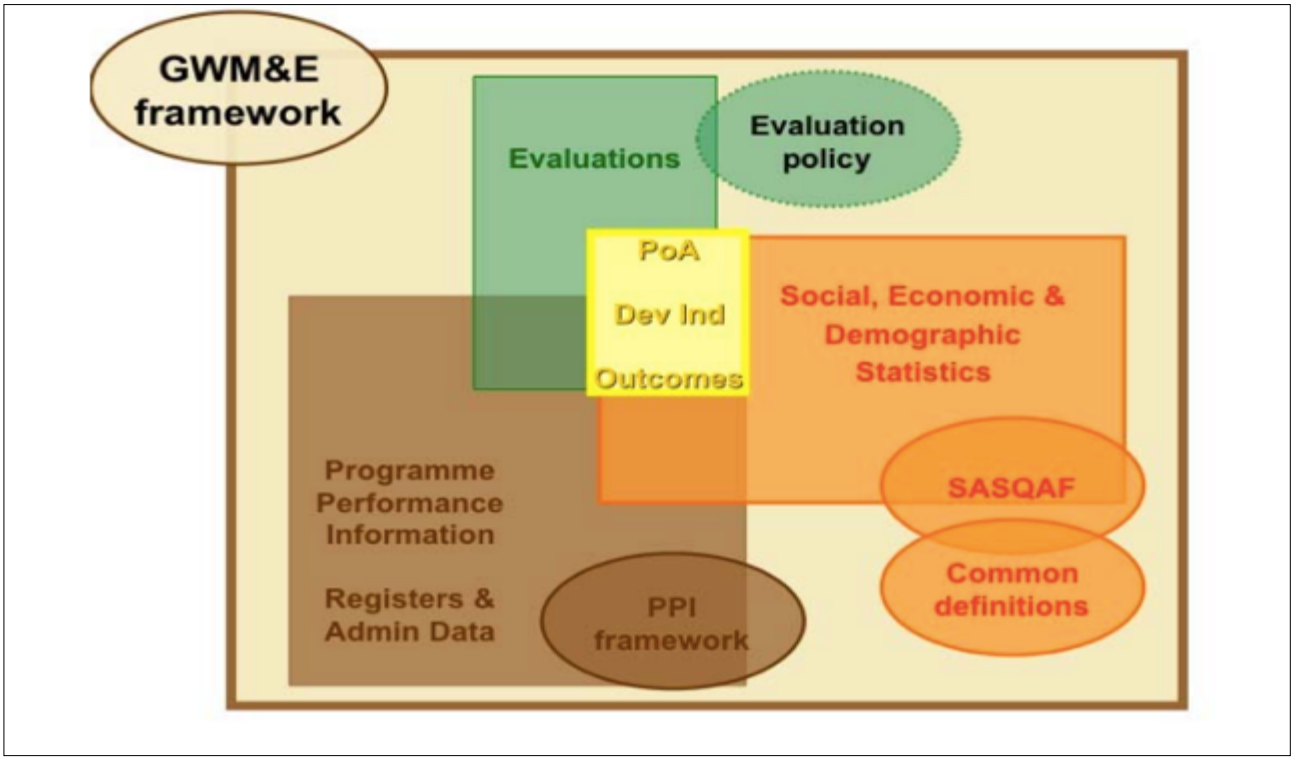


Figure 2.3: The South African Government-Wide Monitoring and Evaluation System

Source: Engela and Ajam, 2010:3.

Oversight over the entire GWMES policy framework is provided by the Presidency. According to Engela and Ajam (2010: 9), the coordination of the GWMES and political championing of government performance management at Presidency level has been critical in strengthening the M&E system. This is in line with international best practice as Lopez-Acevedo, Krause and Mackay (2012:29) point out that some of the most successful M&E systems in the world have a "powerful champion, an influential minister or senior official who is able to lead the push to institutionalize M&E, to persuade colleagues about its importance, and to allocate significant resources to creating a government-wide M&E system". Phillips et al. (2014) confirmed that this strategy has been adopted in the South African public sector with the country's M&E system supported by a designated ministry that demonstrate leadership on M&E and foster intergovernmental collaboration and commitment on government wide monitoring and evaluation.

Components of the system are implemented by key line departments. The National Treasury leads the implementation of programme performance information as prescribed by the Framework for Programme Performance Information, FPPI (RSA, 2007). Statistics South Africa leads the coordination of statistical data as prescribed by the South African Statistics Quality Framework, SASQAF (StatsSA, 2010). Evaluation in the country are prescribed in the evaluation policy called the National Evaluation Policy Framework, NEPF, (RSA, 2011a). Oversight over this evaluation policy is provided by the Department of Planning, Monitoring and Evaluation (DPME).

The NEPF prescribe the establishment of a National Evaluation System that implement and provide oversight over public sector evaluations. As the NEPF has been under implementation since 2011, the National System of Evaluations (NES) is still nascent. With an increasing number of evaluations added to the list every year, this study comes at an opportune time to compare and attempt to synthesise the focus of the evaluations completed under the NES. Government is developing this system through a growing pipeline of planned evaluations that include strategic programmes of government. These evaluations are submitted by all arms of government and form annual National Evaluation Plans (NEP). These plans are updated, approved and implemented annually. This then informs a continuous pipeline of rigorous evaluations in order to inform key decision-making regarding their implementation and impact. Impact evaluations are one of the evaluation types prescribed in NEPF and therefore one of the key evaluation types to feature in the National Evaluation System (NES).

The NEP, has been developing since 2011. Whilst not fully mature, it currently has 54 evaluations that are completed and in progress (DPME, 2016: ix). Appendix A includes all the evaluations that are in the NEP from 2011 to 2015. Current approved evaluations are shown in Appendix B and proposed evaluations are shown in Appendix C.

Annually, national departments propose evaluations, which are reviewed by a working group for inclusion into the NEP, after which they are funded by both the DPME and the submitting department. The final NEP receives Cabinet approval for its implementation. All completed evaluation reports are approved at Cabinet level and made publicly available after which improvement plans and implementation of evaluations are monitored (Porter & Goldman, 2013: 7).

In synthesising the focus of the current completed evaluations listed in the NEP portal, it becomes evident that there are less than a handful of completed ***impact*** evaluations. An analysis of current approved evaluations in the NEP for the 2016/17 financial year listed evaluations that are predominantly implementation evaluations. Therefore, a challenge is presented because, whilst the National Evaluation Policy Framework identifies impact evaluations as critical, few are being conducted and completed as seen in past evaluation plans, current plans and proposed plans for 2017/18, all listed in Appendix A, B and C.

This could be an indication of where the South African public sector is in terms of its M&E trajectory. Implementation evaluation interrogates the integrity of the implementation chain in terms of enhancing effectiveness. In addition, impact evaluations require solid data emanating from good programme implementation. Therefore, poor programme implementation and erratic or incomplete M&E data negatively impact the feasibility of future impact evaluations of programmes. Since the South African public sector is to a large extent embedding M&E systems, impact evaluation of programmes may be premature.

Research by Porter and Goldman (2013: 7) found that monitoring dominates the M&E system in South African public sector, and there remains issues of focus between results and outcomes orientations including the quality of the data emanating from the M&E system. Therefore, the current monitoring and evaluation environment across government in South Africa has largely been one of measuring activities, rather than the effective measurement of outcomes. This has resulted in malicious compliance reporting and focus on internal monitoring rather than focus on outcomes.

It is also argued that, for the reason that impact evaluations demand a greater level of skills and expertise, intense resource allocation, and are far more comprehensive and in-depth in nature, they are not easy to conduct. This has implications for the commissioning of impact evaluations as issues of quality and standards come to the fore and if these are not evident then fewer impact evaluations are commissioned. Porter and Goldman (2013: 8) highlighted this issue, arguing that “development of evaluation norms and standards can help government to place demands on the evaluation profession that will raise the overall quality of practice. Further local capacity can receive preference in commissioning evaluation, rather than relying upon international expertise”.

Consequently, this research found that there is a paucity of **completed impact evaluations** within the National Evaluation System. Whilst some impact evaluations are planned for in the annual evaluations plans, few are actually implemented and concluded. A review of the status of evaluations as at September 2015, reflected in Appendix A, illustrated that in the 2012/13 fiscal year, a total of 8 evaluations were in the NEP. Two of these were impact evaluations, one was successfully completed and the other one was stopped. The following year 2013/14, a total of 16 evaluations were in the NEP. Five of these were impact evaluations, most of which were at draft report stage and were planned to be tabled at Cabinet for approval. In the 2014/15 fiscal year, a total of 15 evaluations were in the NEP. Five of these are termed ‘impact/implementation’ evaluations. Progress on these was varied with some at draft report stage, one evaluation stopped, other two evaluations at service provider selection stage and one evaluation approved by Cabinet. Finally, in the 2015/16 fiscal year, 11 evaluations are planned for in the NEP. Out of this total, only one evaluation is termed an ‘impact/implementation’ evaluation. This particular evaluation was also not implemented due to ‘insufficient budget.

A summary of approved evaluation for the 2016/17 fiscal year as reflected in Appendix B, indicates that, nine evaluations are planned for of which only one is an impact evaluation. Appendix C reflects proposed evaluation for the 2017/18 fiscal year. In this summary of proposed evaluation, there is no evidence of proposed impact evaluation as yet.

Therefore, this analysis clearly demonstrated that impact evaluations are not widely represented within the various annual NEPs. This has posed a challenge for this research due to the small sample population. Additionally, the sparse number of impact evaluations conducted annually is a shortcoming for the NES as these are one of the evaluation types prescribed in NEPF and therefore one of the key evaluation types to feature in the NES. This is specifically relevant when considering the importance of the *government outcomes approach*. In this context, the evaluation of government programmes, implemented and supported with public money, is essential in order to meet the requirements of public accountability and transparency as an integral part of good governance. Therefore, outcome accountability through the strengthening of the national evaluation system is essential. As pointed out by Phillips et al. (2014: 396) and Engela and Ajam (2010) the monitoring and evaluation of key and strategic programmes of government is critical to ensure that outcomes rather than activities remain the focus of government. Therefore, achieving the 14 outcomes of government as detailed in the government' s strategic plan, the Medium-Term Strategic Framework (MTSF) 2014-2019 (RSA, 2014) require greater probity and verifiable evidence. By so doing, "opportunities for political 'spinning' in the system and giving the illusion of accountability" can be averted (Engela & Ajam, 2010: 15). In addition, the National Development Plan (NDP) emphasised that such demands for accountability will broaden the achievement of long term better outcomes in policy areas such as education, healthcare, sustainable human settlements and youth unemployment amongst others (RSA, 2012).

This limited number of impact evaluation in the NES also possibly indicates the capacity and capability challenges of conducting these types of evaluations, as impact evaluations are arguably the most theoretically rigorous and resource intensive type of evaluations. It also indicates that key focus is on programme implementation issues, rather than on impact and attributions, as the data systems that inform that type of analysis are still in development. Porter and Goldman (2013: 8) argued "a common challenge is that impact evaluation of programmes is desired, but this has not been designed from the outset (so a counterfactual is a challenge). Consequently, innovative methodologies are needed, the skills for which may be lacking". This research aims to contribute towards enhanced impact evaluation in light of the identified challenges and observed gaps.

2.9 SUMMARY AND CONCLUSION TO CHAPTER 2

In the preceding analysis, an overview of the existing literature, both recent and prior, on the nature of evaluation was illustrated. Evidence-based policy-making provided the contextual factors and rationale for the emergence of the knowledge generation policy nexus. An attempt was made to locate Realist Evaluation within the wider evidence-based policy debate as key contributor in the systematic review of evidence.

Emanating from that analysis, the review unpacked the various methodological schools of thought about the nature of evaluation. The 'methods branch' of evaluation thinkers was found to have been influenced by the experimental methods applied in natural sciences. The experimental methods were interrogated in the following section to understand the evaluation results emanating from such methods and the evidenced limitations in terms of answering the question of why and how the treatment worked or not was probed.

Ensuing from that discussion, the review provided an overview of global programme impact evaluation current theory and practice. This led to a discussion on the current M&E environment in South Africa, within the overarching context of the results-oriented public sector. The current context of M&E in the South African public sector was discussed as influenced by the overarching Policy Framework of Government-wide Monitoring and Evaluation (GWMES), its oversight, implementation and key components. A key component the National Evaluation Policy Framework, (NEPF) was unpacked including its prescriptions for the establishment of a National Evaluation System (NES) that implement and provide oversight over public sector evaluations. The content and structure of the implementation tools of the NES, the annual National Evaluation Plans (NEP), that form the pipeline of planned evaluations of strategic government programmes were reviewed. It was highlighted that there is a paucity of completed impact evaluations within the NES. Whilst some impact evaluations are planned for in the annual evaluations plans, few are actually implemented and concluded.

A key critical review of the current theoretical positions and practice in both the global and local South African context is that these have been heavily influenced by the traditional positivist research paradigm. In the South African context, based on the evaluation theoretical methods and positions adopted and adapted, the Euro/Western positivist approach has been implemented uncritically and taken for granted as the prevailing conventional wisdom. Alternatives and other theoretical frameworks informed by lived experiences of South Africans, local constructs and contextual factors have not been fully appreciated. In this regard the role of evaluation research in bringing about evolution and social change has been underestimated by evaluators and commissioners alike.

According to House & Howe (2000) social justice is achieved through deliberate democratic evaluation that includes all stakeholders, engages in dialogue with them to uncover their needs and interests as well as deliberate upon such preferences and values. The reviewed methodological approaches lack such broader tentacles as the methods sought to find what works. Future research in a South African context could possibly expand and encompass these social justice and democratisation aspects which propel evaluation research to contribute to social change.

A key theoretical contribution of this chapter has been a synthesis of the evaluation foci of the National Evaluations completed to date. The study finds a severe shortage in commissioned and completed impact evaluation studies which provides insights on the current environment and the impact on the National Evaluation System. As impact evaluations often address the core question

of the policy-maker, namely ‘did it work?’, addressing the current lack of impact evaluations will become a crucial challenge to overcome in expanding and furthering the aims of the NEPF.

Following from this discussion, the next chapter progresses the discussion to the emergence of Realist Evaluation Method with the examples of application from the international context.

CHAPTER 3:
THE EMERGENCE OF THE REALIST EVALUATION METHOD

3.1 INTRODUCTION

The aim of the review in this chapter is to gain an in-depth understanding of the Realist Evaluation Method (REM) through a detailed review and analysis of the related literature as well as to compare international experiences of the use of Realist Evaluation methodological approach in conducting impact evaluations in the public sector.

This chapter begins by providing the current context of the methodological approaches in impact evaluations globally. Within this backdrop, the emergence of Realist Evaluation Method is presented, including the theoretical foundations of the approach, the key ideas of realist inquiry and its research application. Its current application in the international public sector is discussed and examples from international application are highlighted.

Emanating from this background, the suitability of Realist Evaluation and ideal settings for its application are discussed including the success of the Realist Evaluation on programme evaluation and its limitations. Emanating from this extensive review, an assessment framework that adopts a Realist perspective that can be applied to assess the limitations of other impact evaluation methods is proposed.

3.2 METHODOLOGICAL APPROACHES IN IMPACT EVALUATION

This research focuses on programme impact evaluation that seeks to ascertain what would have occurred without the intervention in place. The literature terms this as the ‘counterfactual’ that seeks to assign attribution from a specific intervention. The idea is to discover what would have happened in the absence of the intervention and whether the observed changes and outcomes could be evident as a result of the implemented intervention. Figure 3.1 below locates this type of evaluation within the programme logical framework.

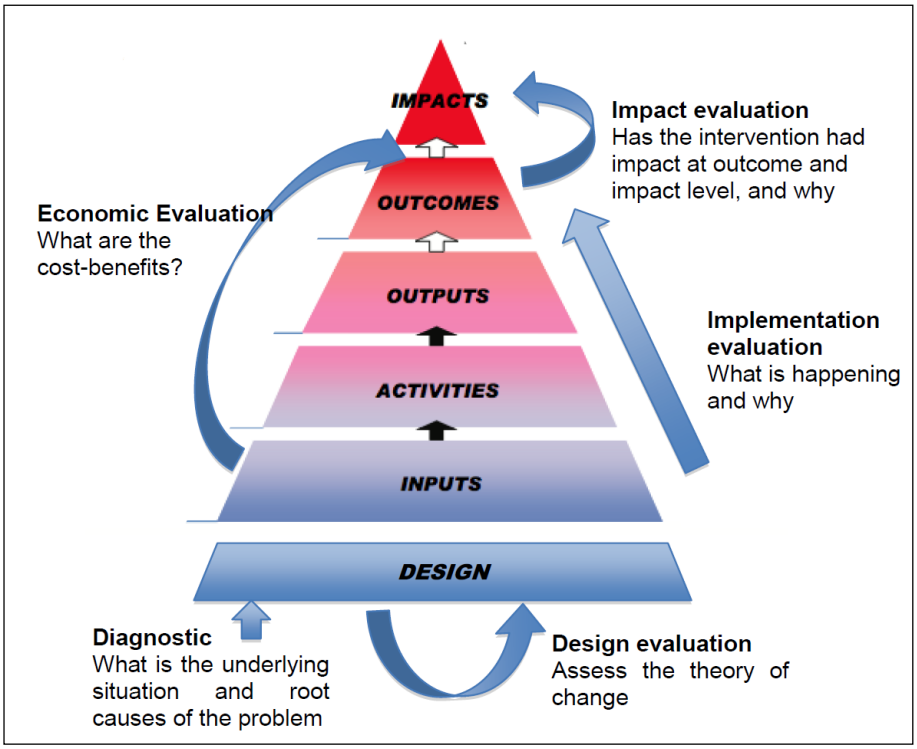


Figure 3.1: Types of evaluations

Source: RSA, 2011a:8.

Traditionally, randomised controlled trials entailing experimental design methodologies were regarded as the gold standard in conducting impact evaluations (Suchman, 1967; Cartwright, 2007 in Pearce & Raman, 2014; Marchal, Westhorp, Wong, Van Belle, Greenhalgh, Kegels & Pawson, 2013: 125). Davies et al. (2000: 263) pointed out that randomised controlled trials are regarded as the 'gold standard' for benchmarking the validity and methodological 'purity' of evaluations.

It has been argued that experiential evidence from RCTs is prioritised over theory with overarching reliance on controlling the experimental context (Davies et al., 2000: 265). Randomised studies have traditionally been conducted in individualistic treatment models, primarily in the healthcare field where the methodology was developed and refined (Suchman, 1967). This later saw the proliferation of randomised programme impact evaluations.

However, in the late 20th century, there was increasing dissatisfaction with the evaluation results emanating from these methods and these were increasingly challenged due to some identified limitations. These limitations entailed the inconclusive nature of experimental design methodologies in answering the question of why an intervention worked in a particular manner. Other researchers (McDonalds, 2000: 120 in Davies et al., 2000) found that there is some scepticism and antipathy on the validity of randomised controlled trials in social evaluations because of their prioritisation in research. Davies et al. (2000: 261) pointed out that the conditions and contextual factors surrounding the intervention, can negatively impact the experimental design and make the entire experiment unsustainable.

Table 3.1: Methodological challenges in randomised evaluation studies

Methodological challenges in randomised evaluation studies	
Ethical concerns	Randomly allocating individuals to different interventions raises serious ethical and sometimes legal issues to do with informed consent.
Learning curves	Many interventions may take time to be perfected. The question then arises as to when evaluations of these should be undertaken. Too soon and even those with potential are likely to be rejected; too late and ethical considerations may preclude randomisation.
Variable delivery	The delivery of many interventions may rely on the skill of the deliverer. Ensuring consistent and replicable delivery may be difficult leading to concerns about what exactly is being evaluated.
Interactions	There may be interactions between the intervention deliverer and the intervention recipient, which affect the likelihood of achieving beneficial outcomes.
Individual preferences	Individuals may have strong prior preferences that make their random allocation to groups problematic or even unethical.
Customised interventions	Some interventions may need considerable customising to individuals to achieve best effect. Concerns again rise as to what is being evaluated.
Lack of group concealment	Knowledge of group allocation may lead to changes in behaviour or attitudes, which undermine the unbiased nature of any evaluation.
Contamination	Understanding by participants of the nature of the evaluation may lead to convergence between the groups diminishing any effect sizes.
Lack of blinding	Blinding is difficult or impossible for many interventions with the attendant risk of bias being introduced by study subjects' behaviour, compensatory activities by external agents, or differential outcome assessment.
Poor compliance	Individuals allocated to one intervention or another may fail to comply with stipulations, thus undermining the assessment

Source: Davies et al., 2000: 262.

There were increasing calls for evaluations that have both internal validity (in terms of experimental design) and external validity (in terms of the generalisability), utilisation value as well as incorporating qualitative elements.

Bamberger (2008: 122) illustrated the limitation of randomised controlled trials and pointed out the flaws inherent in these methods since the:

> *…decision to use randomized control trials, or strong quasi-experimental designs, can affect the programme being evaluated. For example, the use of randomization means that selected beneficiaries will include both poorer and better-off families from the target population. However, if beneficiaries were selected by the implementing agency, it is possible that preference would have been given to the poorest families.*

Therefore, those included in many experimental trials are not characteristic of the sample populations, nor are they necessarily suitable for addition in the relevant study. Selection centred on chance may not target the correct envisaged beneficiaries.

Pure experiments are meant to eliminate bias in all respects in order to ensure that any observed changes are as a result of the administered intervention. Proponents of RCTs believe that the internal validity and rigour of RCTs are crucial to counter-act the selection bias that is inherent in non-randomised evaluations. On the other hand, critics have pointed out that though RCT-based evaluations have strong control in their design, they lack external generalisability to participants outside the study. Evaluation results and actual impacts resulting from these studies have been questioned for their lack of utility value and inconclusive nature. The inconclusive nature of evaluations conducted through RCTs resulted in the under-utilisation of evaluation findings. Bamberger (2008: 134) suggested that poor understanding of the applied evaluation methodology may inadvertently lead to lack of utilisation of the evaluation findings.

Patton (1996: 65; 2012: 364) on the other hand, indicated that evaluation findings only have use value to stakeholders on condition that the primary purposes of conducting the evaluation are addressed. The findings need to enable the rendering of judgements about the extent to which the bulk of the evidence lead to significant findings and conclusions as to the observable impact of the intervention including, facilitating of improvements and generating new knowledge. If any of these purposes are not identified and utilised by the evaluation stakeholder group, then the evaluation findings failed to be appreciated or utilised, resulting in minimal or no impact in policy-making. Consequently, a successful evaluation should be informed by an appropriate methodology that is sound and acceptable to key stakeholders.

Weiss (2004a: 154) noted that “evaluation is a big tent, and there is room within it for evaluators with diverse perspectives”, whilst Chen (2005: 416-421) further posited that there are many evaluation approaches available. Evaluation practitioners should have a selection menu of appropriate methods to cater for the evaluation needs at hand. Given that other impact evaluation methods, such as experimental methods, have shown limitation in enlightening policy-makers about how or why impact occurred or failed, the exploration of a theory-based evaluation approach in conducting programme impact evaluation in the South African public sector is well timed and appropriate.

The discussion document *Improving Government Performance: Our approach* issued by the Presidency (RSA, 2009: 21-22) recommended the review and strengthening of theory-based evaluation approaches and articulated that:

> *While many public sector organisations have developed comprehensive indicator sets as part of their planning and reporting processes, analysis of causal effects is currently weak, and the international good practice of theory-based evaluation needs to be strengthened. This would require that in the policy development and planning stages a clear conceptual understanding of how, why and when the policy, programme or project will effect change, and how these changes may be measured.*

The South African public sector acknowledges that on programme impact evaluation the analysis of causal outcomes is not strong enough and there is merit in exploring theory-based evaluation approaches. In addition, Chen (1994: 231) has made clarion call for more research in this area, saying:

> *…there are still not enough research examples for demonstrating how to apply various theory-driven evaluations for different evaluation needs and situations. For a greater usage of theory-driven evaluation, there is an urgent need in the immediate future to develop prototypes for a variety of theory-driven evaluations.*

One such theory-based evaluation approach is the Realist Evaluation Method (REM) (Chen, 1990; Pawson & Tilley, 1997; Pawson, 2013). The utilisation of experimental methods such as RCTs as the gold standard in conducting programme evaluations has been long standing. The REM offers an alternative to the experimental approach in programme impact evaluation and therefore its potential value in the public sector necessitates further analysis. Realist Evaluation offers causal analysis of programme efficacy by unearthing the evidence of 'what works, for whom, in what context and in what respects'. Pawson (2006: 25) emphasised that in Realist Evaluation the three elements of context, mechanism and outcome must be considered in order to answer the question of 'what works'. The REM is an emerging programme evaluation method that is based on evidence, based on articulation of programme theory and is explanatory in providing insight on programme effectiveness; i.e. the missing 'why' factor that is absent in other methods. The REM does this by not merely asking whether the programme works or not, but rather by searching for evidence on 'what works, for whom, in what context and in what respects'.

In addition to REM, two other impact evaluation methods, Outcome Mapping and Contribution Analysis contribute certain elements and components that are found in Realist Evaluation. Outcome Mapping is another impact evaluation method that seek to uncover changes or outcomes in the behaviours, relationships, actions or activities of those interacting directly with the programme, known as boundary partners (Earl, Carden & Smutylo, 2001: 1). Outcome mapping focuses less on the programme's outputs and the changes on conditions as a result of the programme, but casts focus on the behaviour of the people impacted by the programme. The behavioural changes that the programme influences on the beneficiaries are the key focus. The similarities between outcome mapping method and REM reside in the interrogation of the behaviour of programme beneficiaries who are regarded as key agents of change. In REM, the behaviour and thought pattern of the intended beneficiaries as they engage with the interventions result in the observed programme mechanism. Similarly, outcome mapping regards the behaviours of the people, groups and organisation that are recipient interventions as the key factor that will capacitate them to build and sustain the outcomes of the programme. Therefore, whilst REM looks at the behaviours of the recipients as a catalyst that results in change mechanism, it goes further than outcome mapping and interrogates the programme context as another factor that contributes to a programme's outcomes.

On the other hand, the contribution analysis method recognises that measuring a programme's contribution to outcomes is not an exact science. Therefore, the method seeks to reduce uncertainty regarding a programme's influence on certain outcomes and rather increase understanding of how that is achieved through a collection of additional data and information that increase understanding about a programme and its impact. This is achieved by interrogating the programmes theory of change to find out what programme pathways to change are (Mayne, 2001: 5-6). The similarities between contribution analysis and REM are a mutual focus on programme theory of change as well an interrogation whether the programme theory is supported by the outcome patterns or the 'performance story'. REM, however, further interrogates the programme mechanism as well as the programme context to gain further insight on programme impact.

This research aims to discover whether the Realist Evaluation Method could be practical and useful in enhancing the impact evaluation of public social programmes in the South African public sector as an additional evaluation method. By so doing, this research aims to contribute towards new knowledge on improved evaluation methods that are evidence-based with valid and robust findings.

This discussion has argued with other impact evaluation methodologies that were considered, such as experimental methods. It has been argued that experimental methods have not demonstrated a strong link between the mechanism of why and how programmes achieve outcomes or not, the impact of contextual factors on programme impact and programme beneficiaries. Pawson and Tilley (1997: 8) maintained that "experimental paradigm constitutes a heroic failure, promising so much and yet ending up in ironic anti-climax". Justification for the choice of the Realist Evaluation approach as a potential method for impact evaluation in the public sector has been provided as well as its selection as the method to be researched.

This methodological approach towards impact evaluation holds the promise of providing a strong evidence of 'what works, for whom, in what context and in what respects' as most public sector programmes are implemented in large, complex, multi-faceted social environments with little or no understanding of causal mechanism. In this vein, Davies et al. (2000: 271) argued that:

> *…robust evidence of what works is needed to inform policy and practice; evidence that is robust both in terms of its internal persuasiveness as well as its external applicability. What is also needed is robust evidence of how things work to allow progress in intervention design and tailoring to specific contexts.*

3.3 DEVELOPMENT AND RATIONALE FOR REALIST EVALUATION

This section discusses the development and rationale for Realist Evaluation emanating from the context of evidence-informed policy-making. The discussion further locates the method within the theory-based approach.

3.3.1 Theory-based evaluation

The previous section presented trends in the methodological approaches used in evaluation. The section highlighted the importance of method-driven evaluations modelled on RCT experimental methods that aimed to preserve methodological purity. Method-driven approaches were influenced by the work of Campbell and Stanley (1963; 1966) as well as Cook and Campbell (1979) who prioritised the importance of internal validity in programme evaluation at the expense of theory, resulting in narrow and limited analysis. Stame (2004: 59) stated that in most evaluations a lot of resources and effort were spent preserving experimental internal control and proving generalisability at the expense of clearly articulating programme theory. As a result, there was a strong focus on being method driven rather than theory driven with the resulting implication that the theory of change of how programmes caused observed change was assumed and not interrogated. Furthermore, the primary aim was to rather verify the internal validity of the programme and this was deemed to imply programme rationality.

The development of a theory of change of an intervention serves to demonstrate explicitly and logically the change mechanism of an intervention. Proponents of theory of change refer to “theory-based evaluation” (Weis, 1995; 2000) and “theory-driven evaluations” (Chen, 1990). Rossi and Freeman (1985; 1993), Weiss (1990; 1997a; 1997b; 1997c; 2000; 2004a; 2004b; 2008) and Chen (1990; 1994), pioneered thought leadership on the specification of the theory of change embedded in programmes and the advocacy of evaluation testing such theories.

Chen’s influential work (1990) claimed that social science knowledge and theory are critical in understanding programme evaluation. However, social science theory has not dealt effectively with programme theory, tending to offer abstract theorising and general propositions in this regard whilst the ‘how to’ aspects have been underestimated Chen (2005:18). According to Chen (1990:43) programme theory is “a specification of what must be done to achieve the desirable goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated.” These implicit and explicit assumption of what it is that must be done to achieve the envisaged outcomes are programme theory. In this regard programme theory encompass both descriptive and prescriptive elements. The descriptive elements, detailing the causal processes, are critical as they specify how the programme is supposed to work. If they are unsound the strategies of the programme will not work. Therefore, the effectiveness of the programme is contingent on the effectiveness and validity of these described causal assumptions or causative theory. Chen (2005:17) called these descriptive

assumptions the ‘change model’ and this is important for developing the justification of the programme as it describes how change happens. The second element of programme theory is the prescriptive or normative element. What should be done to achieve the envisaged outcomes. These prescriptions prescribe what should be done to activate the change model. (Chen 2005:18). This ‘normative theory’, ‘prescriptive theory’ or ‘action model’ of the program dictate what implemented components and activities will be required to activate the change model (Chen 2005:18).The explicit definition of the action model, so far as it provides guidance and directions to evaluators on what should be done to effectively implement the programme, is important. Hence, an understanding of both the change and action models is crucial for the effective implementation programme policy. In this regard the action model offers the ‘nuts and bolts’ of how to assemble the resources required to action the change model of programme. Poorly defined change and action models in programme theory can result in programme failure that is primarily due to inadequately defined causal linkages and a level of poor implementation. Consequently, whilst social science theory can explain programme theory at a conceptual level, an understanding of the practicalities of programme implementation are important.

Chen has further advocated conducting evaluation using both social science knowledge and traditional research methods such as randomised experiments, arguing that, given the experimental and quasi-experimental dominant paradigm, inconclusive results from such experiments do not adequately explain programme failure. Therefore, these should be supplemented by models drawing from programme theory. In this regard, Chen and Rossi (1983: 300) argued that:

> *…we have argued for a paradigm that accepts experiments and quasi-experiments as dominant research designs, but that emphasizes that these devices should be used in conjunction with a priori knowledge and theory to build models of the treatment process and implementation system to produce evaluations that are more efficient and that yield more information about how to achieve desired effects.*

Therefore, the experimental paradigm with its lack of theory may provide inconclusive results and sometimes inaccurate understandings of programmes. Accordingly, theory-driven evaluation enriches experimental methods and may lead to the highly sought-after dual attainment of both internal validity and external generalisability, as well as studying the causation link within the context of constructed theories.

Chen and Rossi (1983: 300) have argued that programme evaluation should:

> *…move from the black box evaluation, which is concerned primarily with the relationship between input and output of a program, to the theory oriented evaluation, which emphasizes an understanding of the transformational relations between treatment and outcomes, as well as contextual factors under which transformation occurs.*

By opening the 'black box' that method-driven evaluation failed to do, theory-based methods provide the 'enlightenment' aspect that is sorely missing in programme evaluations. According to Chen and Rossi (1983: 299), most evaluations that have been traditionally conducted have "been social accounting studies that enumerate clients, describe programme and sometimes count outcomes". Pawson and Tilley (2007: 4) further supported this view by proposing that present day evaluation faces a predicament since there is a large burgeoning of data, lack of accumulation of results and research that remains lacklustre and inconclusive with little or no impact on policy-making.

The testing of the programme theory entails identifying the integral aspect and mechanism that causes the observed outcomes. This entails finding out causal attribution, through a mix of experimental and quasi-experimental and programme theory. Weiss (2000: 35) pointed out that "theory-based evaluation tests the links between what programmes assume their activities are accomplishing and what actually happens at each small step along the way". Theory-based evaluation, which uses an explicit theory of change concept, starts with a linear line of thought or causal chain and describes in a schematic presentation how the intervention is projected to generate results and outcome. The theory of change explains the sequence of actions and events that will lead to the expected results attributable to the intervention. The overarching programme theory of change further describes the mechanism of how the programme results in change, including describing the assumption, the risks and the ideal context that will result in those expected results. Where programme theory is complex, there will be multiple theories as to how a programme works with concomitant varied causal links. The evaluation will inform as to which links and which theories are investigated, since investigating all possible theories and all causal links may be improbable. In deciding which theories to select and which links to study, Weiss (2000: 44) advised that the underlying mechanism on which the programme rests must be considered with due consideration for the plausibility of the conjectured theories and practicalities of access, resources and methodological capabilities of the evaluation.

Theory-based methods serve to answer the questions that are in the 'black box' of experimental methods, since the causal links between the intervention and observed results are interrogated. Rogers, Petrosino, Huebner and Hacsi (2000: 5) described theory-based evaluation as "an explicit theory or model of how the programme causes the intended or observed outcomes and an evaluation that is at least partly guided by this model". Thus, theory-based evaluation examines and confirms the programme causal model. Rogers (2007: 7) argued that, in order to conduct theory-based evaluation, programme theory must initially be defined whether by evaluators or stakeholders closer to the programme. Thereafter the defined programme theory informs the evaluation approach. The main thrust of theory testing in theory-based evaluation has been the testing of programme impact through assigning causal attribution. This can highlight issues of theory failure and implementation failure. Chen (1990: 57) cautioned that programme theory is value driven on 'ideas or meanings of what ought to be' and these values can result in different programme theories constructed from both

the stakeholders' and evaluators' perspectives. To counter-act this conflicting value judgement in generation of programme theory, Chen (1990: 60) suggested programme evaluation should have fundamental values of stakeholder responsiveness as advocated by Stake (1976), Patton (1984; 1986; 1990; 1996; 2008; 2012), Cronbach (1980) and Lincoln and Guba (1986). Scriven (1999; 2003/2004) has called for evaluator objectivity in evaluation and Chen (1990:62) supported this view and advocated that objectivity is critical as 'evaluators' personal preference and/or loyalties do not contaminate evaluation results'.

Chen (1990: 63) further advocated that trustworthiness is a critical value in the construction programme theory and described trustworthiness as an evaluation providing 'convincing evidence that can be trusted by stakeholders and others in the utilization of evaluation results'. In this view, the work of Campbell and Stanley (1963) and Cook and Campbell (1979) on the importance of internal validity provides a basis for this value. Finally, Chen (1990:65) stressed the important value of generalisability, which is "the extent to which evaluation results can be applied to future pertinent circumstance or problems to which stakeholders are interested". This future-oriented approach of programme evaluation was influenced by the work of Cronbach's (1982) generalisability theory on future value and utilisation of evaluation by decision-makers.

3.3.2 Theories of change

Theory-based evaluations may be regarded as a new wave of programme evaluation moving away from the quantitative and qualitative paradigm that has been the hallmark of the evaluation. The move beyond methodical paralysis is deemed to enable better programme analysis. In this context, Rogers (2007: 63) pointed to the proliferation of terminology that has defined programme theory including labels such as "theory-based", "theory-driven", "theory-oriented", "theory-anchored", "theory-of-change", "intervention theory, "outcomes hierarchies", "programme theory", and "programme logic".

Stame (2004: 60) emphasised that this new wave ensures that programme theory is central during conceptualisation of an evaluation and all methodical approaches are considered and valued for their suitability and input. In this context, all methods are complementary, since programme theories drawing from both qualitative and quantitative methods enable better articulation of the programme's causal chain and key assumptions. This view is strongly supported by Pawson (2002a: 158) who claimed that "evaluation research has come to understand that there is no one 'gold standard' method for evaluating single social programmes".

The use of programme theory is now commonplace and it has been mainstreamed and phenomenally applied in programme management processes in various programme areas. Programme theory is a core requirement as evaluation commissioners require project proposals to initially specify the theory of change as a guide for assisting in programme design and evaluation (Rogers, 2007: 63-64). The theory of change is validated and tested by verifying the extent to which the theory assumptions are true on what is actually observed. Theory-based evaluation as opposed

to method-driven evaluation approaches provides for more rigorous evaluation through systematic interrogation of programme theory as a basis for guiding the evaluation.

However, much of what is termed ‘programme theory’ in contemporary evaluation practices is simply superficial implementation theory as depicted by the linear ‘logic model’ specifying programme inputs, activities, outputs, outcomes, and impacts. Rogers (2007: 64) argued that these schematic representations are inadequate in explaining the full programme as issues of causality are not articulated. The absence of causal mechanism in logic models results in limited information on how implemented programmes can be implemented in others settings. Therefore, linear logic models are an acceptable first step in unpacking programme theory, but the process should go further than a simple logic model to truly represent programme theory with causal mechanism.

Programme theory is sourced from various stakeholders and is not simply informed by the evaluators’ opinion. Rogers argued (2007: 65) that best practice in programme theory evaluation includes the interrogation of causality, application of current research theories, and complementary methods in order to crystallise programmes mechanism. This is best achieved by the engagement of all relevant stakeholders, as well as the testing of programme theories sourced from all programme stakeholders and through a process of competitive elaboration and testing against the data. This is the same view as that of Pawson and Tilley (2004: 10-11) who argued that the programme theory’s generative mechanism demands some research skills in sourcing the information from “documents, programme architects, practitioners, previous evaluation studies and social science literature” as necessary. The theory can be developed at the programme’s inception by the programme architects and implementers or after the programme is underway (Rogers et al., 2000: 7). In some instances, where there is no explicit programme theory, the evaluator can develop the theory based on review of related programme literature to ascertain the programme causal mechanism informed by similar programmes, discussion relevant programme stakeholders, through review of programme documentation, or through observed evidence (Lipsey & Pollard, 1989 in Rogers et al., 2000: 7).

According to Weiss (1997c: 78), the future direction of programme evaluation based on analysis of programme theory is to limit the theory analysis to the dominant programme postulations with a view to crystallise key drivers of programme success and these should be contrasted with other similar programmes. Therefore, the focused testing and validation of theory can yield consistent evidence about the validity of the underlying programme theory. In this regard, Weiss (2000:39-41) specifies the critical elements to be considered in the formulation of programme theory Weiss. First the theory assumptions should be sourced from all stakeholders as specified by Rogers and Pawson. Secondly, the chosen theory must pass the plausibility test in terms of the reasonableness of the theory in comparison to the observed outcomes during implementation. Thirdly, the theory that presuppose that the provision of additional information to the target group result in changed behaviour should be tested in order to determine how additional information to programme

participants undermines or enhances programme outcomes. Lastly, a relevant programme theory will be critical to the programme's success as the programme will largely be a translation of this theory.

In conclusion, Realist Evaluation and theory of change approaches have become prominent and popular methods (Blamey & Mackenzie, 2007: 439). Weiss (2000: 44) conceded that, due to the astronomical variety of programme implementation, variety of stakeholders, structural factors, socio-political contexts and available financial resources, it may be a challenge for evaluation findings to meet the external generalisability threshold. For this reason, theory-based evaluation can add even relatively incremental knowledge about programme mechanism. From the ensuing analysis, it is evident that Realist Evaluation is a leaf of the theory-based evaluation branch. Since the overarching aims of theory-based evaluation are to open the 'black box' of the experimental paradigm and offer enlightenment as to how and why programmes work, Realist Evaluation as an affiliate of this method, could possibly enhance understanding of how programmes work. Evidence indicates that experimental paradigms overemphasise methodological purity and internal validity with demonstrated lack of attention to programme theory. These limitations inevitably provide inconclusive results and sometimes inaccurate understandings of how programmes work. Accordingly, Realist Evaluation enhances the possibility of achieving both internal validity, external generalisability as well as the understanding of causal links within the context of constructed theories.

3.3.3 An overview of the Realist Evaluation approach

The aim of this analysis is to gain an in-depth understanding of the REM approach underpinned by its philosophical approach with the aim of defining the key ideas of realist inquiry and its research application. The epistemology, ontology and social theory of Realist Evaluation is rooted in European philosophical traditions (Pawson & Tilley, 1995: 20). In philosophical Realism, what is real is deemed to exist independent of the mind and therefore reality exists independent of human thoughts and beliefs. Realism's epistemology is based on scientific positivism, which adopts a systematic approach to the generation of knowledge. What is seen is a representation of reality. Such an external reality exists that can be discovered by means of a systematic, interactive methodological approach (Auriacombe, 2013: 725). Greenhalgh et al. (2015: 3) claimed that this reality represents "a real social world but that our knowledge of it is amassed and interpreted, sometimes partially and/or imperfectly via our senses and brains, filtered through our language, culture and experience". Hence, interventions can operate and result in varied programme mechanisms based on the contextual factors at play.

In the philosophy of science Realism offers the third way after empirical positivism and relativism (Sayer,2000:2). According to the Realism ontology, the world is a highly complex, structured and open system made up of people, institutions and societal structures. These societal structures, organisations and institutions shape and define human behaviour and can exist without the actors

deliberately shaping them. The complexity that is embedded in the social system, influence human behaviour, who act in specific ways as defined by their context and circumstances. Theory in these circumstances seeks to explain the resulting outcomes from such social interactions. This entail understanding the implicit causal mechanism that result in the observed outcomes. Bhaskar (2008:3) claims that "generative mechanism is nothing other than the ways of acting of things. And causal laws must be analysed as their tendencies, which may be regarded as powers and or liabilities of a thing which may be exercised without being manifest in any particular outcome." Empirical research, in whatever form, must then be assembled to provide evidence of how a specific causal mechanism comes about. Conjectures are then made about how the causal mechanism work and this is compared to the observed outcomes to provide an explanation.

Within the Realism epistemology, two dominant ontologies have become prominent, Critical Realism and Scientific Realism, also referred as Empirical Realism. According to Sayer (2000:11), Critical Realism should not be confused with Empirical Realism which is akin to empiricism and identifies the real world with the empirical. Furthermore, Archer, Bhaskar, Collier, Lawson & Norrie (1998: xi) claim that Critical Realism is able to combine and reconcile ontological realism, epistemological relativism and judgmental rationality. Bhaskar (2008: xxix), a theorist of this ontology, pronounced that Critical Realism provide "a comprehensive alternative to the positivism that has usurped the title of science". He therefore proposes 'transcendal realism' or Critical Realism as an alternative. This 'transcendental critique' of positivism is from the basis that the results of laboratory experiments are not causal laws since such observed regularities emanate from an artificially closed environment and cannot be perpetuated outside such an environment. In social science, attempts at approximating laboratory conditions whilst studying social science phenomena has led to a debate on whether social science should be a critical exercise or an empirical science. There is equally a dilemma in social sciences on whether society is an open complex system or whether it can be construed as closed system where phenomena can be experimented on under closed conditions (Pawson, 2006:18). In this regard, the transcendental method, refutes empiricism and transcendental idealism (Bhaskar, 2008: xi). It does this through an immanent critique of the positivist interpretation of science. This is a crucial departure point between Critical Realism and Scientific Realism.

Scientific Realism view the positivist empirical methods that recognise both the observable and unobservable phenomena relevant in explaining phenomena. Though the methods of science maybe imperfect, such as the closed system, science makes progress in explaining the unknown. These methods can be strengthened by Scientific Realism whose key features are its explanatory focus and these explanatory strategies can inform a gradual body of scientific knowledge (Pawson & Tilley 1997b: 406). According to Pawson (2006:17), Critical Realism, veering from the original aims of scientific explanation as constructed by the forerunners of Realism, assumed a normative critical moral lens to critically evaluate human activity. Critical Realists contend that this immanent critique

is relevant as it refutes empiricism and transcendental idealism (Bhaskar, 2008: xi). Scientific Realism therefore distance itself from Bhaskararian Critical Realism as that branch is deemed to be normative and righteous. This is construed as misaligned with the aims of Realist Evaluation which seek to advance evidence base policy making. Therefore, Realist Evaluation is implicitly aligned with Scientific Realism or Empirical Realism.

Research by Auriacombe (2013: 725) has found a variety of evaluation approaches alternating from a positivistic, deductive perspective, followed by a pragmatic approach that is both deductive and inductive, as well as a post-modern inductive and constructivist paradigm. Within these paradigms, Realist Evaluation was located within the pragmatic paradigm. This view is supported by Greenhalgh et al. (2015: 3), who claimed that realism can be thought of as sitting between positivism ('there is a real external world which we can come to know directly through experiment and observation') and constructivism ('given that all we can know has been interpreted through human senses and the human brain, we cannot know for sure what is the nature of reality').

In contrast, Pawson, Wong and Owen (2011) claimed that Realist Evaluation is "Popperian and Campbellian in its philosophy of science and thus relishes the use of the brave conjecture and the application of judgement". The philosophy of science from Hesse (1974), Lakatos (1970), Bhaskar (1975) amongst others influenced realist synthesis (Pawson & Tilley, 1997a: 55; Pawson & Tilley, 1997b: 405). According to Pawson (2013: 1-12), the pillars of realist wisdom rest on giants such as Bhaskar's (1978) generative mechanism; Archer's (1995) individual choices and resulting social outcomes; Elster's (2007) theories on explaining social behaviour; Merton's (1967) middle range theories that deal with different spheres of social behaviours and structures; Popper's (1992) theories on the link between scientific discovery and cumulative evidence; Campbell's (1988) call for objectivity in evidence and competence at both quantitative and qualitative levels; and Rossi's (1987) observation of opposing mechanisms in large-scale social programmes. The philosophical wisdom of these thinkers influenced the emergence of the Realist Evaluation Method as propounded by Pawson and Tilley's seminal work (1997a). Realist Evaluation and the realist synthesis are research strategies that represent the principles of realist explanation and philosophy of scientific realism.

Pawson (2013: xix) pointed out that Realist Evaluation's standpoint is that of pursuing the high scientific objectives of objectivity and generative causal explanation and inform real world policy and practice. Figure 3.2 provides an overview of Realist Evaluation.

Realist Evaluation's epistemology is founded on theory that provides theory of change explanations and such explanations are continuously refined and regenerated based on emerging programme evidence (Pawson & Tilley, 1997a; Pawson, 2006; Pawson, 2013).

Realist Evaluation is self-confessedly theory driven, investigates, and enhances the underlying reasons of programme effectiveness (Pawson, 2013: 15; Pawson & Manzano-Santaella, 2012: 178).

Its overarching aims are the systematic discovery of how circumstantial influences stimulate the connection between the intervention and outcome. Pawson and Tilley (1997a: 56) added that its hallmarks "are its explanatory mechanisms and its attempt to show that the usage of such explanatory strategies can lead to a progressive body of scientific knowledge".

Central to the method is the conceptual framework of context-mechanism-outcome configuration or CMO (Pawson & Tilley, 1997a). The proposition is that programmes have successful outcomes through a unique combination of the underlying thought processes and actions of programme participants (the mechanism) who act in appropriate contextual environments. Successful and effective programme impact evaluation is a result of deciphering a suitable context and the underlying mechanism that results in planned outcomes. According to Pawson and Tilley (1995: 23), "the success of programs will be highly conditional. Things work if the circumstances are right. Effects occur only if the conditions are right, and they may have to be very particular".

Therefore, Realist Evaluation seeks to understand the 'generative causation' or 'generative mechanism' of a programme in order to understand how and why change occurs. In this regard, Pawson and Tilley (1995: 29) specified that:

> *realists conceive of causality in generative terms. Thus, instead of Y simply following X and that being the beginning and end of what can be said, realists consider the causal powers or causal potential that inherent in phenomena and that may be released in some circumstances to produce observable transformations.*

Therefore, the right conducive conditions will generate causality or set off a trigger or 'fire' that generate the observed outcomes. This mechanism is at the heart of Realist Evaluation and seeks to understand "the reasoning processes of those implicated in change (or non-change) and [and] why people do or do not decide to behave in one way or another need to be built into the evaluation" (Pawson & Tilley 1995: 23).

The context-mechanism-outcome configurations (CMO) is a proposition that specify under what context and how will the outcomes be achieved. This is because programmes do not offer a single theory of how they might work. Each inquiry offers various mechanisms of how a programme can work, within various contexts and circumstances of the actors involved, resulting in various outcomes.

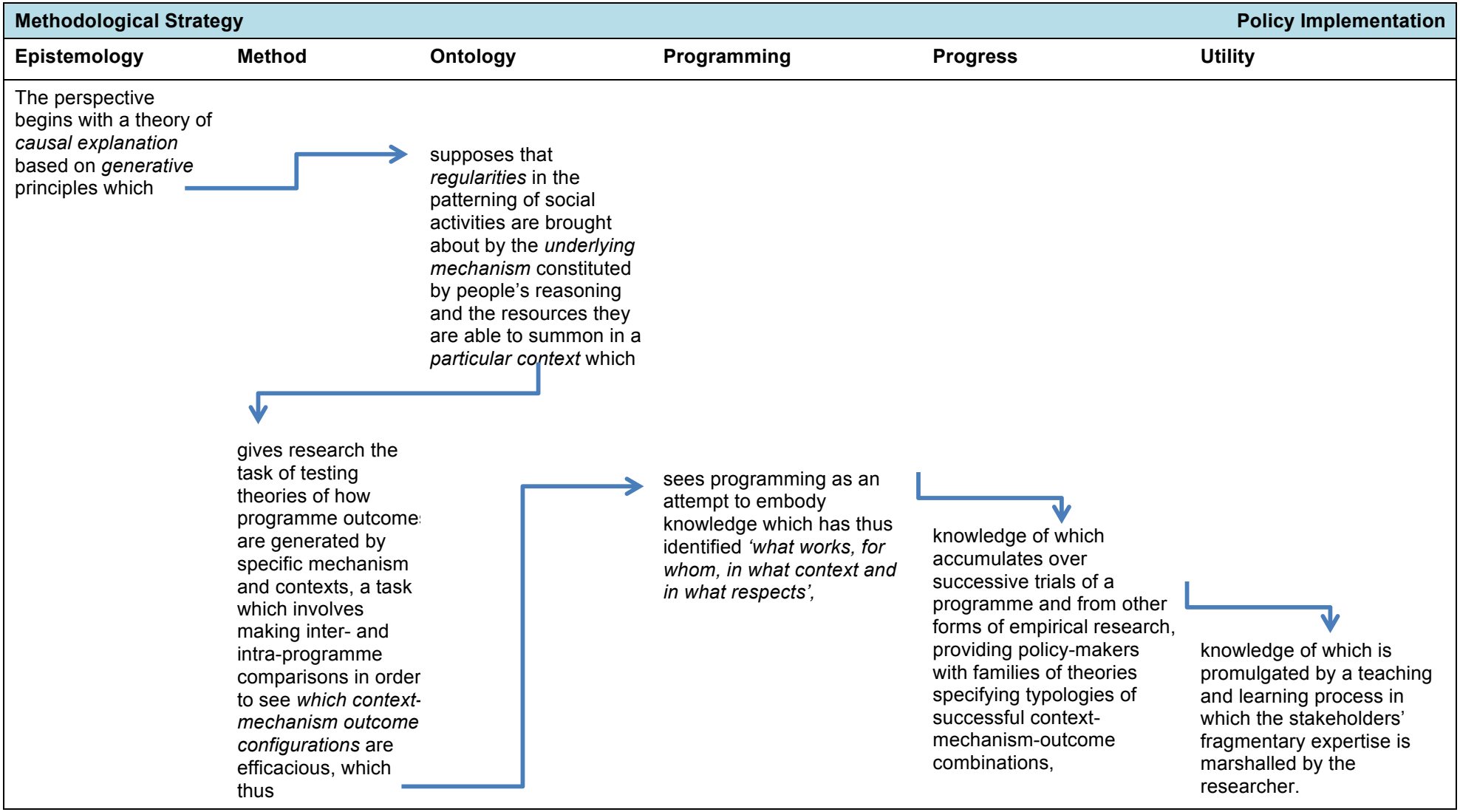


Figure 3.2: An overview of Realist Evaluation

Source: Pawson and Tilley, 1997a: 220.

The CMO table illustrated below in Table 3.2 is a visual way of integrating a list of propositions within any programme evaluation. According to Pawson and Manzano-Santaella (2012:183-184), CMO are configurations and specific propositions that draw specific contexts, with specific mechanisms and outcomes and these are best illustrated in a table format as an '*if then*' proposition and are read horizontally.

Table 3.2: Multiple CMO configuration propositions

Context	+	Mechanism	=	Outcomes
C_1	+	$\mathbf{M_1}$	=	$\mathbf{O_1}$
C_2	+	$\mathbf{M_2}$	=	$\mathbf{O_2}$
C_3	+	$\mathbf{M_3}$	=	$\mathbf{O_3}$
C_N	+	$\mathbf{M_N}$	=	$\mathbf{O_N}$

Source: Adapted from Pawson, 2013:23.

$C_1M_1O_1$ presents the elemental proposition of how the programme works. This initial configuration is put to test, investigated and produces specific middle-range theories pertaining to C_1 as illustrated in Figure 3.3 below.

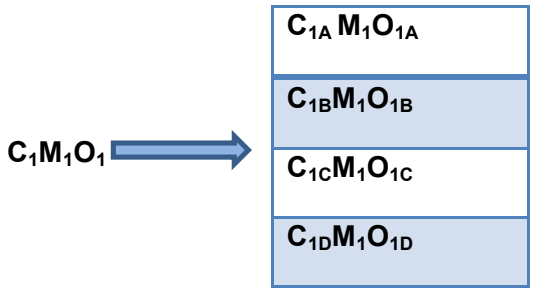


Figure 3.3: CMO before and after testing

Source: Adapted from Pawson, 2013:22.

$C_2M_2O_2$ can also be tested and it produces its specific, differing conditionalities, $C_{2x}\, M_2O_{2x}$. The process continues with subsequent CMOs tested and producing various $C_{Nx}\, M_N\, O_{NX}$ hypotheses. The objective of testing is to understand programme mechanism. Pawson (2013: 23) pointed out that multiple propositions for testing should be made based on the available time and financial resources. Programmes have endless complexity, embedded in complex situations; change can be engendered in myriad ways, driven by various actors responding in various ways.

Therefore, Realist Evaluation develops and tests various conjectured CMO configurations empirically. The overarching strategy is to test the adequacy of the theory in different contextual settings. Through this testing and refinement of theory, the research finding will present an empirically-tested theory of change which explains causality.

According to Pawson (2013: 14), an evaluation is not applying the realist research technique if there is an “absence of an explanatory focus; working in one data medium method, rather than being multi-method; and failure to investigate contexts, mechanism, and outcomes in configuration”. Realist research design is versatile and pluralist as the research methods are varied and choice of method is determined by the hypothesis.

Pawson and Tilley (1997a: 114) pointed out that the Realist Evaluation technique is a research logic, guided by the three themes of Realist Evaluation strategy. These themes of Realist Evaluation strategy increase the specificity of our understanding of the mechanism through which a programme accomplishes change; intensify the understanding of programme context that influence programme mechanism; and expand the preciseness of forecasting programme outcomes. This “realistic explanatory triad” attempts to open the ‘black box’ of programme mechanism.

Realist Evaluation findings are intended to assist in the effective tailoring of programmes. Realist Evaluations do not query, ‘What works?’ or ‘Does this programme work?’ but rather ask, ‘What works for whom in what circumstances and in what respects, and how?’ This is precisely because “programmes never work indefinitely, in the same way, or in all circumstances, nor do they work for all people” (Pawson & Tilley, 1997a).

Pawson and Tilley (1997a) argued that the context under which a programme is implemented can affect the outcome. Therefore, it is critical to understand the underlying causal mechanism of a programme in order to shed light on how a specific outcome was achieved. In realist tradition, there is no assumption that interventions in and of themselves bring about change instead the way interventions are delivered and the contexts that surround them are the factors that trigger the causal relationships. Realist Evaluation views programmes as being rooted in the societal structures and systems, and the effectiveness of the programmes is influenced by the context under which they are implemented. Programmes are sophisticated social exchanges operating amidst a complex social reality. Since programmes are the theories of policy-makers or programme designers, the entire effectiveness of programmes is dependent on how well the programme theories were crafted. The context under which a programme is implemented is critical as the context can be supportive or unhelpful in the success of the programme. This context can be social, economic and political structures including institutional contexts, the programme staff and participants, the geographical area where the programme is implemented and historical context.

Contextual factors combined with programme participants’ reasoning all result in a specific mechanism of how the programme may actually work and this ultimately results in observed outcome patterns. Certain factors in the context trigger or inhibit particular mechanisms influenced by the unique combination of the context and mechanism. Therefore, Realist Evaluation will seek to test each context-mechanism-outcome (CMO) configuration of the programme until arriving at the most plausible outcome based on evidence.

Consequently, Realist Evaluation pursues the comprehension of 'what works, for whom, in what context and in what respects' in order to replicate a programme effectively. Therefore, programmes in and of themselves do not work, but it is rather the opportunities offered to programme participants that trigger certain responses resulting in programmes that work. Rather than offering prescriptive findings, Realist Evaluation offers policy options accompanied by limitations and deliberations that should advise the decision-making process. Realist Evaluation recommendations come from systematic evidence that commences with proposed programme theory, which undergoes testing and concludes with programme theory that has been refined through empirical evidence.

In the programme evaluation sphere, there is ''a serious shortage of rigorous, systematic evidence that can guide evaluation or those evaluators can use for self-reflection or for improving their next evaluation" (Henry & Mark, 2003: 69). Whilst in the current public policy sphere policy-making based on solid evidence is valued, the application of evidence practically remains a challenge due to the complex nature of programmes (Pawson, 2002b: 340-358). Consequently, there is emergent attention paid to the extent of theory-driven, qualitative and mixed-method approaches to provide a stronger evidence base of 'what works'. "Systematic approaches as these suggest the prospects of expanding the evidence base by providing the causality of complex interventions" (Greenhalgh et al., 2011).

Pawson (1997; 2013) has claimed that Realist Evaluation is essentially multi-method, which enables the in-depth interrogation of programme impacts. The approach can be applied in all types of programme and policy evaluations. The overarching aims of the approach is to produce empirically-tested programme theory to shed light on how programmes work. The systematic nature of Realist Evaluation lends itself to the implementation of valid, sound and effective programmes. Contextual factors are critical in Realist Evaluation analysis since it may be a possible misnomer to assume that the deemed success of a particular government programme can be replicated indiscriminately in various environments. A programme can succeed in one environment, but the very same programme executed in exactly the same way may dismally fail in another environment. Realist Evaluation seeks to address this assumption and stresses that causality is dependent on a system being activated under conducive conditions. The Realist Evaluation Method is based on the proposition that the same problem existing in various environments cannot be solved by one panacea. Indiscriminate replication of programmes without asking the key Realist Evaluation questions in the evaluation process carry the risk of superficial generalisation in terms of conclusions about programme efficacy.

In summary, Pawson and Tilley's (1997a) seminal work proposed that social programmes are theories of change designed by policy-makers to address existing social problems. These programmes only succeed in achieving the intended outcomes if the programme participants are provided with better choices. Programme participants will assess the rationality of what is offered by the programme and make a choice. Programme participants are reasonable actors, whose reasoning is influenced by their contextual environments, their beliefs and attitudes, and the

available resources on offer. The context under which a programme is implemented, and the combination of the choices made by the participants in that context, influenced by their reasoning and the offered programme resources, the triggered mechanism, are what will make a programme work and create the programme outcome and ultimate impact. The fundamental insight for policy-makers to achieve programme success is to empirically understand the right context, the desired mechanism to be triggered to achieve the intended programme outcome and then to proceed to transfer and replicate the programme in those exact contextual environments as proven by the refined programme theory.

3.3.4 The application of the research technique in Realist Evaluation

Realist Evaluation is similar to the scientific research cycle. In empirical research the logic of inquiry is framed by a theory that explains the abstract. A hypothesis derived from the theory is stated. Through empirical evidence the hypothesis is tested. The result of that observation informs the generalisation confirming the hypothesis or rendering the hypothesis null. Realist research design is distinct from this cycle essentially on the type of data that is collected (Pawson & Tilley, 1997a: 84).

Realist Evaluation begins by understanding the programme theory of change, data is collected to test the programme theory, then various context-mechanism-outcome pattern configurations are tested against the gathered data. From these systematic analyses, there will be diverse emerging outcome patterns detailing what has worked and what has not within the programme itself, and in comparison to other programmes. The research technique entails the interrogation of outcome patterns by searching and looking for associations and contrast between the sub-groups making up the programme. Programme explanations will emerge and a process of adjudicating the plausibility of these explanations will unfold until there is one plausible emerging programme theory, which is then assessed and interpreted against the original programme theory. The overarching aim is to develop, test and refine programme theory. Understanding the initial programme theory is critical in Realist Evaluation.

Pawson (2013: 15) asserted that “Realist Evaluation is avowedly theory-driven; it searches for and refines explanation of programme effectiveness”. The theory of change of Realist Evaluation is based on understanding the logic inquiry of context-mechanism-outcome configuration. This ‘realist explanatory triad’ attempts to lighten the ‘black box’ of programme mechanism or how change occurs. Pawson (2013: 21-22) emphasised that a “CMOC is a hypothesis that the programme works (O) because the action of some underlying mechanism (M) which only comes into operation in particular context (C)”. Therefore, programme theories are conjectures which are casted as if-then propositions, which are then tested to prove that, if the correct methods are in place under conducive conditions, then the expected outcomes will occur.

As illustrated in Figure 3.2, Realist Evaluation follows the conventional positivist research cycle. It begins with the Realist Evaluation proposing that a causal outcome follows from mechanism acting in context. A hypothesis is made about what might work for whom and in what circumstance for a particular intervention. Therefore, the programme theories are elicited, formalised and tested. Research, both quantitative and qualitative, is collected to determine whether the hypothesised Contexts (C), Mechanism (M) and Outcomes (O) are true or not. Consequently, the Contexts (C), Mechanism (M) and Outcomes (O) configuration is confirmed and refined resulting in a researched programme specification of what it is about the programme that produces results, for whom, and under what conditions. A programme can have variations in the configuration of the CMOs. Realist evaluations empirically test each of the CMO's configuration and result in findings that propose types of CMOs that can work in sustaining the programme and result in the achievement of programme outcomes. Therefore, programme implementers are not simply presented with findings relating to cause and effect or unqualified generalisations as is the case with other approaches, but rather presented with a specific menu of CMOs that are ideal for the success of the programme. Realist Evaluation entails the following four stages, discussed in detail below:

- Theory and hypothesis;
- Data collection;
- Data analysis; and
- Theory testing and refinement.

3.3.4.1 Stage 1: Theory and hypothesis

According to Pawson (2004:10-11), "Realist Evaluation normally begins by eliciting and formalising the programme theories to be tested in CMO terms and what is involved is bringing the imagination to bear in 'thinking through' how a programme works". The objective in this stage is to have a clear understanding of the basic initial programme theory of change. Then various hypotheses of CMOs for potential testing in table format are elicited through workshops with various stakeholders and programme source documents. These conjured hypotheses should at best meet the purpose and evaluation question to be answered.

Westhorp (2014:10) suggested asking the following questions in order to elicit programme theory:

- "For whom will this basic programme theory work and not work, and why?
- In what contexts will this programme theory work and not work, and why?
- What are the main mechanisms by which we expect this programme theory to work?
- If this programme theory works, what outcomes will we see?"

In addition, a hypothesis grid (Pawson & Tilley, 2005: 27), illustrated below in Table 3.3, is useful in listing all conceivable and probable mechanisms, contexts and likely outcomes.

Table 3.3: The realist hypothesis grid

Some plausible mechanisms	Some potential contexts	Some possible outcomes
M_1	**C_1**	**O_1**
M_2	**C_2**	**O_2**
M_3	**C_3**	**O_3**
M_4	**C_4**	**O_4**

Source: Pawson and Tilley, 2004:27.

Therefore, the initial hypotheses are made on what it is about the programme or intervention that works, i.e. the "generative mechanism", "for whom, under what circumstances will it work", how will it work and why. Pawson and Tilley (2004: 10-11) urged that good research skills in sourcing the information from "documents, programme architects, practitioners, previous evaluation studies and social science literature" are necessary. The information is articulated and prioritised as theory propositions or hypotheses to be empirically tested.

Figure 3.4 below illustrates this stage and the subsequent stages of Realist Evaluation cycle.

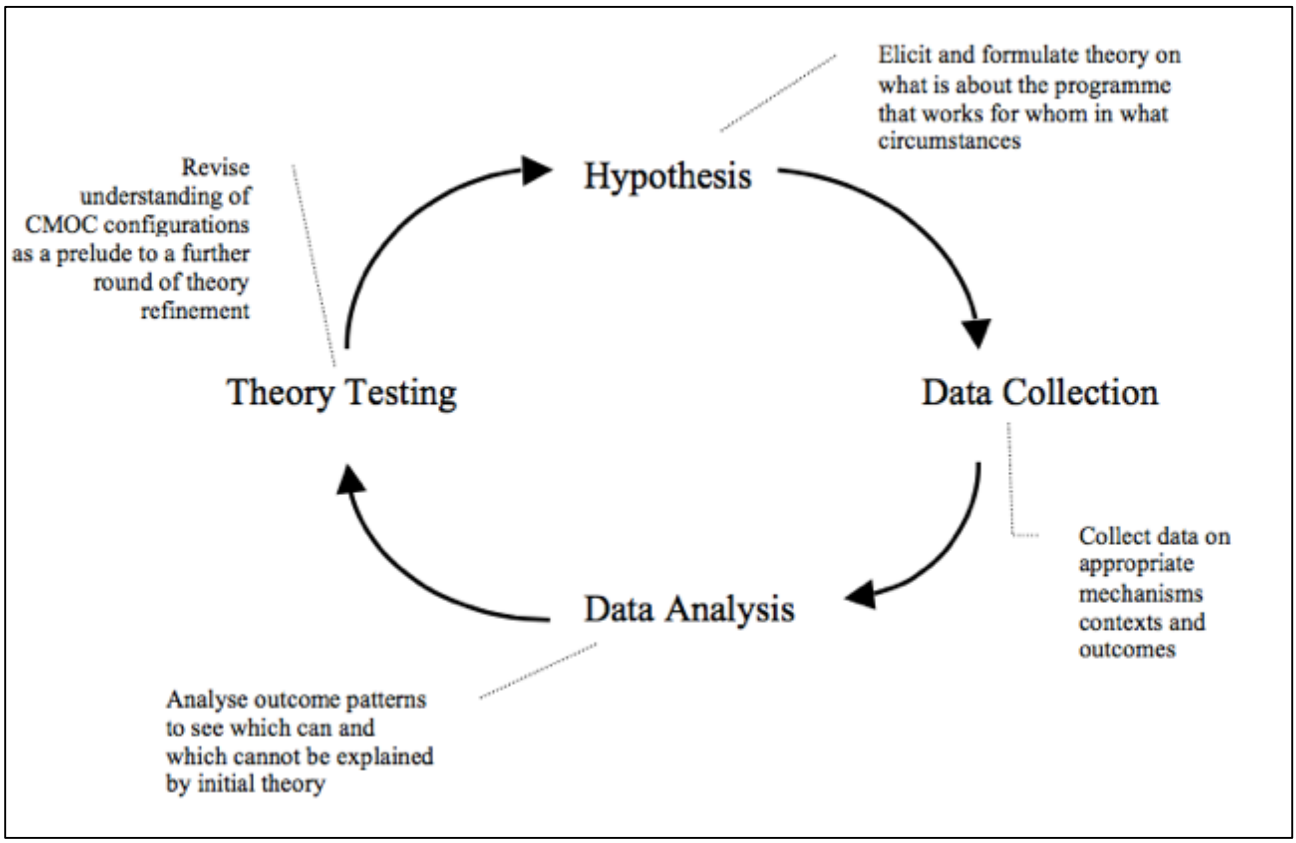


Figure 3.4: The Realist Evaluation cycle

Source: Pawson and Tilley 2004:24.

3.3.4.2 Stage 2: Data collection

Pawson and Tilley (2004: 11) claimed that this stage entails "collecting data that will allow interrogation of these embryonic hypotheses". The specified programme goals inform the collection of appropriate data on mechanisms, contexts and outcomes. Both qualitative and quantitative data is collected to ensure adequate testing of theory. The collection of data and the adopted research methods should be informed by the evaluation question at hand in order to test the theory and effectively answer the evaluation question.

3.3.4.3 Stage 3: Data analysis

The third stage according to Pawson and Tilley (2004: 11). "is to subject a whole package of CMOC hypotheses to systematic tests, using data sets assembled in stage 2". The primary objective of this stage is to search for evidence of the emerging outcome from the ensuing collected data. Various "context-mechanism-outcome pattern configurations theories" are tested against the gathered data. The data analysis aims to test the initial theory against the observed outcome patterns of the programme.

Marshal, Van Belle, Van Olmen, Hoerée and Kegels (2012: 195) described this stage as entailing:

> *…analysis of qualitative data from interview transcripts and documents based on coding in terms of 'description of the actual intervention', 'observed outcomes', 'context conditions' and 'underlying mechanisms'. Quantitative data is analysed with the aim of assessing the effectiveness of the intervention and to substantiate or invalidate the patterns that emerge. The resulting explanations for the observed outcomes are formulated as refined conjectural CMO configurations.*

These conjectured CMO configurations can be presented in the form of succinct storylines and graphic presentation to facilitate comparison. Pawson and Tilley (2004: 11) indicated that the observed outcome pattern will have varied degrees of success and failures at intra-programme and inter-programme level. The key strategy is to probe the emergent outcome patterns through iterative critical review of rival explanations regarding programme outcomes juxtaposed against the initial programme theory. Advancing and arbitrating between competing explanations for programme outcomes are informed by research evidence. The overarching aim is to develop, test and refine programme theory. Marchal et al. (2012: 195) added that "through triangulation, the plausible patterns or demi-regularities that explain how the intervention led to the observed results are confirmed". This enables the ascertainment of the programme mechanism and the delineation of who the programme is optimally suitable for.

3.3.4.4 Stage 4. Theory testing and refinement

Pawson and Tilley (2004: 11) asserted that "the final stage is the assessment and interpretation of the analysis. Have the theories about how the programme worked been supported or refuted by the proceeding analysis?"

Based on the research findings, the CMO configurations can be revised and be subjected to a further round of theory testing until the results of the analysis explain the observed variations in outcome patterns. Marchal, Dedzo and Kegels (2010: 3) supported this view emphasising that "realist study ends by adapting the initial theory accordingly. This modified theory then serves as a new hypothesis of the next study. This cycle refines the theory and leads to better insights of how particular interventions work, in which conditions and how".

Pawson and Tilley (2004: 11) confirmed this and stated that the last phase is an iterative and continuous process that aims to ascertain and explain the specifics of programme mechanism. Realist Evaluation empirically tests each of the CMO's configurations and results in findings that propose the types of CMOs that can work in sustaining the programme resulting in the achievement of programme outcomes. Therefore, programme implementers are not simply presented with findings relating to cause and effect or unqualified generalisations as is the case with other approaches, but rather presented with a specific menu of CMOs that are ideal for the success of the programme.

Therefore, the key principles of Realist Evaluation as summarised by Mouton (2007: 507) are:

- "Programmes are theories incarnate;
- Evaluations are tests of theories;
- Theories need to comprise context-mechanism-outcome conjectures;
- Mechanisms refer to the ways in which effects are brought about;
- Contexts refer to the conditions for the operation of a mechanism and they are seldom closed;
- Outcomes refer to the effects of a mechanism activated in context;
- Mechanism generally (though not always) involves reasoning and resources;
- Programmes work differently amongst different subgroups."

Therefore, Realist Evaluation supposes that programmes are designed and made alive emanating from the conceptual thinking of policy-makers hence they are "incarnate" theories. They are conjectures of policy-makers who deduce that "if we present these people with these resources (material, social or cognitive) it may influence them to change their behaviour" (Pawson, 2002c: 213). An evaluation is essentially an empirical test of those specified theories. Furthermore, these programme theories must specify clear suppositions about the nature of programme context, context which is influenced by various socio-economic factors, the mechanism of how the programme causes change, change which is greatly influenced by the thinking and reasoning of programme

beneficiaries, as well as the resulting expected outcomes from the implementation of such a programme. Based on all these factors implemented, programmes will work in different ways and result in different outcomes for different beneficiaries under different contexts and circumstances.

In addition, Pawson and Manzano-Santaella (2012: 177) emphasised that the hallmarks of a Realist Evaluation Method must provide an explanation of why a programme works, can apply and test data using various methods and provide an explanation in which contexts a programme may work, how it may work and the envisaged outcomes from such implementation. Within this theoretical context, it is critical to understand the nature of how programmes work in an open and complex social system.

3.3.5 The nature of social programmes

This section analyses and reviews the myriad ways and nuances of how social programmes work using the Realist Evaluation lens. Pawson (2006: 26-34) provided detailed scrutiny of the working of social programmes and how they should be reviewed.

3.3.5.1 Programmes are theories

The implementation of social programmes is based on hypothesis theory. According to Pawson (2006: 26), programmes have propositions that say "If we deliver a programme in this way or we manage services like so, then it will bring about some improved outcome". Case study analysis should search explicitly for the evidence of these propositions and conjectures.

3.3.5.2 Stakeholder reasoning

Change is triggered by the reasoning of the intended stakeholders of the intervention. If the stakeholders respond in an unintended manner, the intervention will not generate the intended change. Therefore, the reasoning of the intended stakeholders must be accurately anticipated in order to influence positive outcomes. Stakeholders include the programme beneficiaries, the programme staff, policy-makers and other actors who provide agency of the programme. If the reasoning and actions of these stakeholders is in sync with the programmes' overarching aims, then the expected change should be observed. In any policy intervention, the "policy architects try to figure out how to get practitioners to deliver the resources, and then practitioners try to figure out how to get programme participants to change their ways, and the subjects try to figure out whether it is worth buying into the deal" (Pawson 2002c: 213).

3.3.5.3 The programme implementation chain

Implementation is essentially the chain of events and agents who implement the programme, from the time it is conceived to the time the programme reaches its subjects and beneficiaries. This implementation chain encompassing the actors and agents of the programme is illustrated in Figure 3.5.

Since there are various actors and agents involved, there is risk of misinterpretation during the implementation resulting in the miscarriage of the overarching aims of the programme and beneficiaries who respond in unintended ways. Pawson (2006: 29) stated that the analysis must "inspect the integrity of the implementation chain, examining which intermediate outputs need to be in place for successful outcomes to occur, as well as noting flows and blockages and points of contention".

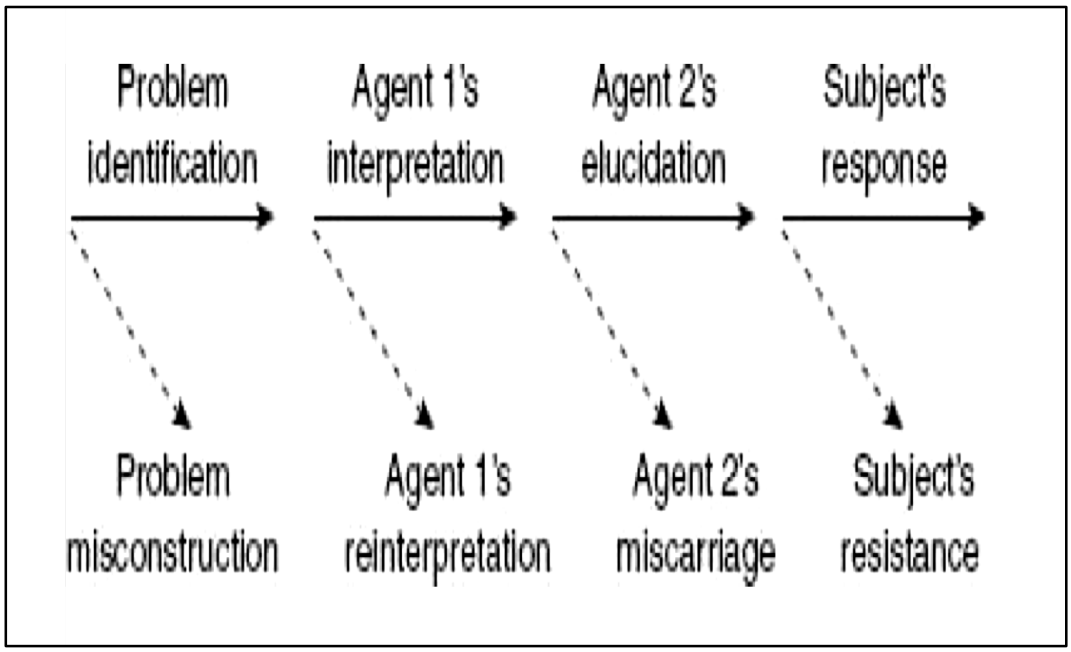


Figure 3.5: Theory chain with intended and unintended outcomes

Source: Pawson, 2006:29.

3.3.5.4 The influence of powerful stakeholders

As the programme is rolled out, some of the agents in the implementation may question some of the assumptions made or the measurement criteria of programme success. Depending on the authority and position of these agents, for instance as subject matter experts, they may seek to influence the redefinition of the initial programme theory or key assumption, essentially resulting in the build-up of programme theory from the bottom up rather that top down. In this vein, Pawson (2006: 30) emphasised that the analysis should "take into consideration how the relative positioning and influence of different decision-makers are able to direct and redirect programme implementation".

3.3.5.5 Social programmes and complex social systems

Social programmes are influenced by their surrounding social environments. The same programme will thrive in one social environment and fail in another setting due to surrounding circumstances and other contextual factors. Pawson (2006: 31) described four factors that shape and influence the context under which social programmes are implemented.

As described in Figure 3.6, these entail: (i) the individual capacities of the key agents and actors who may or may not have the necessary enthusiasm, will and credibility of enabling the implementation of the social programme; (ii) the interpersonal relationship between all programme key stakeholders; (iii) the institutional setting of the implementing agency as defined by its organisational culture and values that serve to enable or inhibit the effective implementation of a social programme; and (iv) the overarching infrastructural system that supports it, such as political backing, resource allocation and positive public perception and support. Therefore, the success of a programme is strongly dependent on its context.

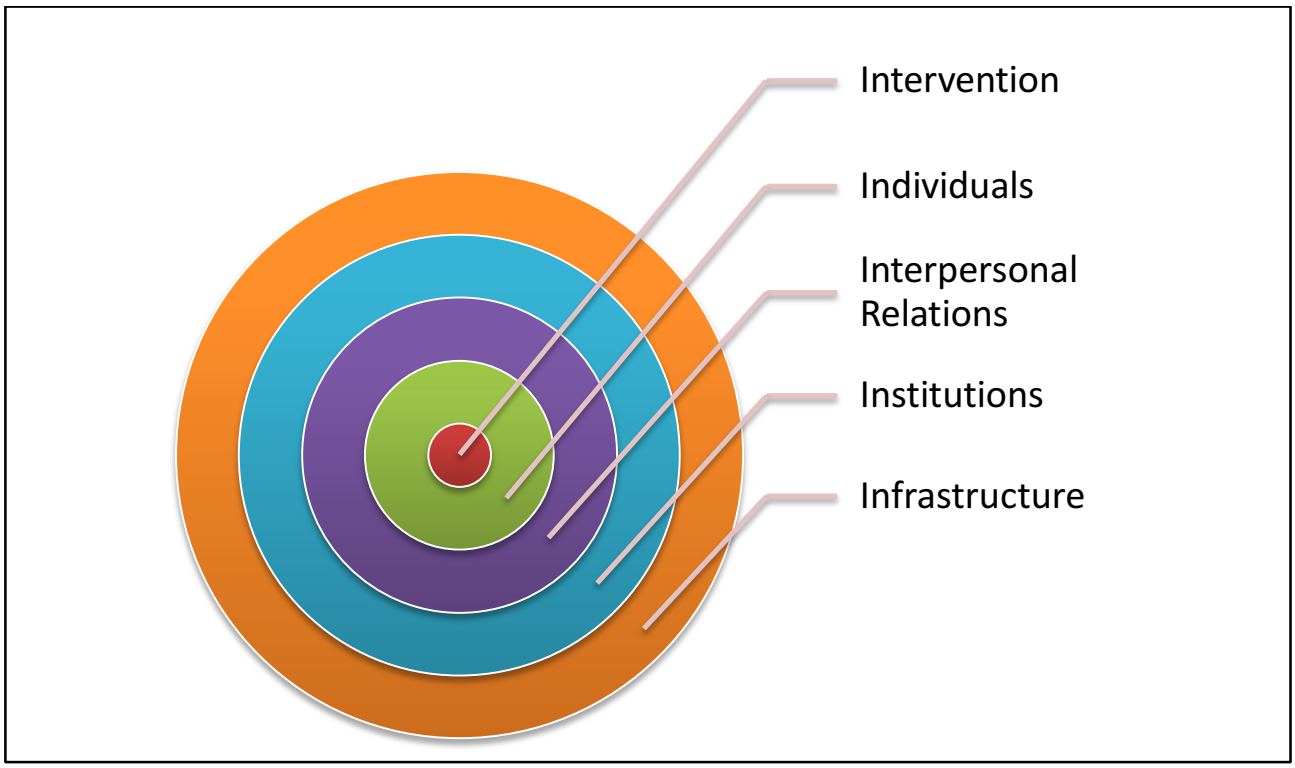


Figure 3.6: The intervention as the product of its context

Source: Pawson, 2006:32.

3.3.5.6 Social programmes as continuously evolving

Through continuous refinement of programme theory informed by emerging theories and evidence from practice, programmes gradually advance from their initial conceptualised theories and evolve and adapt into newer reformed programmes. This is informed by the experiences of practitioners and implementers as they share knowledge and best practice from across programmes. This cross-fertilisation of tacit knowledge results in programme theory that evolves over time.

3.3.5.7 Revitalisation of social programmes

As programmes mature their *modus operandi* becomes fairly standard and known. This can result in loopholes and unanticipated programme manipulation by the various implementation stakeholders. Therefore, social programmes must be reviewed at certain intervals, infused with new initiatives and ideas in order to circumvent change that is impacted by malicious compliance and fraudulent activities. In this regard, Pawson (2006: 34) emphasised that impact evaluation should assess the effects of "familiarization and habituation", that develop as programmes become established.

3.3.6 The application of Realist Evaluation in the international public sector

Internationally, there has been a number of Realist Evaluation applications. Hayton (2015) applied it in UK public sector economic development agencies and Marchal et al. (2012) explored a model for the use of REM within public healthcare. Other research found the application of Realist Evaluation in project-based policy environments (Davis, 2005:275). Others such as Pedersen and Rieper (2008) applied the method to large-scale public sector reform of the Danish electricity sector. Evaluations based on theory-based methods such as Realist Evaluation are increasingly accepted and used.

The examples below demonstrate that there is an emerging body of scholarly research on this approach. Some international applications include cases such as:

- Reflections from a Realist Evaluation in progress: scaling ladders and stitching theory (Punton, Vogel & Lloyd, 2016);
- Aid effectiveness and governance reforms: applying realist principles to a complex synthesis across varied cases (Betts, 2013);
- Large-system transformation in Healthcare: A Realist Review (Best, Greenhalgh, Lewis, Saul, Carroll & Bitz, 2012);
- Realistic Evaluation as a new way to design and evaluate occupational safety interventions (Pedersen, Nielsen & Kines, 2012);
- Turning around an ailing district hospital: A realist evaluation of strategic changes at Ho Municipal Hospitals (Ghana) (Marchal, Dedzo & Kegels, 2010);
- Evaluating the evidence: A move to more Realistic Evaluation: A case study of regional selective assistance in Scotland (Hayton, 2015);
- A Realistic Evaluation: The case of protocol-based care (Rycroft-Malone, Fontenla, Bick & Seers, 2010);
- Is Realist Evaluation a realistic approach for complex reforms? (Pedersen & Rieper, 2008);
- How do you modernize a health service? a Realist Evaluation of whole-scale transformation in London (Greenhalgh, Humphrey, Hughes, Macfarlane, Butler & Pawson, 2009).

Whilst the Realist Evaluation approach is applicable across all policy environments, it has been prominent in healthcare interventions. According to Weiss (1997c: 71-72), much application has been in areas of healthcare due to the fact that healthcare interventions are implemented through well-articulated programme planning processes, procedures and theory of change assumptions, which are well suited to the methods of theory-based approaches. Therefore, through its systematic nature and explanatory focus, Realist Evaluation findings have been found to be value-adding in the effective tailoring of healthcare interventions.

This view is supported by Greenhalgh et al. (2015: 1-2) who found that Realist Evaluation has proved to be conducive to health service research because interventions addressing wellness are complex with multiple causes, both at individual level and in society at large. These include interventions that target smoking, obesity, alcoholism and other social ills. In these varied contexts, interventions do not work uniformly and RCTs and meta-analysis of those trials have provided inconclusive outcomes as to what works. Health research had found Realist Evaluation to advance knowledge and understanding of various healthcare issues. Marchal et al. (2012: 208) supported these views and indicated that Realist Evaluation philosophies have become prevalent in healthcare research as they have been found to be useful in simplifying the complex nature of such interventions. Porter and O'Halloran (2012: 19) concluded that the effectiveness of intervention in healthcare environments is impacted by the open surrounding complex social system that affects the efficacy of the intervention, whilst RCTs neglect this important aspect in their quest to preserve the internal validity of the experiment resulting in a closed system that ignores the social system.

Research by Greenhalgh et al. (2015: 8) has highlighted that, while Realist Evaluation is promising for building resilient programme theory and feeding into the evidence-informed policy-making sphere in various research areas, there remains poor application and misinterpretations of the method. Research by Marchal et al. (2012) and Pawson and Manzano-Santaella (2012) demonstrated that some ostensibly 'Realist Evaluations' were in actual fact misinterpreting the method and not applying the theory underpinnings of the method appropriately and consequently produced findings and recommendations that were flawed. They found that some researchers failed to understand the explanatory nature of the CMO configurations and instead produced laundry lists of contexts, mechanisms and outcomes, which were 'unconfigured' or not properly structured as interconnected relationships. This has highlighted the need for quality standards and criteria for validating true and accurate Realist Evaluation. There are currently no professional guidelines or standards for conducting Realist Evaluations. Having said that, it is noted that the "Realist and Meta-Narrative Evidence Synthesis: Evolving Standards (RAMESES I)" protocol series of projects is under implementation (Greenhalgh et al., 2011). These are protocols aimed at providing guidance and quality standards pertaining to Realist Evaluation and realist systematic reviews.

Another study, “Protocol – the RAMESES II study: developing guidance and reporting standards for Realist Evaluation (RAMESES II), follows in the footsteps of RAMESES I, which was a protocol aimed at producing procedural advice, publication benchmarks and training and coaching resources for the conduct of systematic reviews from the Realist Evaluation perspective (Greenhalgh et al., 2011: 1).

According to Greenhalgh et al. (2015: 4), at present RAMESES II is in the process of developing and issuing the quality reporting standards for Realist Evaluation. The key aims of the project (Greenhalgh et al., 2015: 4) are:

> *… to develop quality standards, reporting guidance and training materials for Realist Evaluation, build capacity for undertaking and critically evaluating Realist Evaluation in the healthcare context and produce resources and training materials for lay participants, and those seeking to involve them, in Realist Evaluations.*

Consequently, there is progress within the international community of practitioners to embark on a path that will provide quality assurance on the application of this approach. The main purposes for developing these standards is to ensure that researchers plan and undertake comprehensive Realist Evaluations as well as provide guidance to research output users on how to judge the validity and dependability of the evaluation outputs (Greenhalgh et al., 2011: 2). By so doing, Realist Evaluation quality standards and reporting benchmarks will be further embedded and solidified within the community of practice.

3.3.7 Suitability of Realist Evaluation

Pawson and Tilley (1997a; 2004) claimed that the strengths of the Realist Evaluation come from the methodological rigour found in the pure sciences and such theory-based rigour enables better interrogation in programme impact and resulting policy-making. Through its programme mechanism analysis and context-based analysis, it can inform the clarification and learning on substantive issues across all policy, practice and organisational boundaries. Pawson and Tilley (2004: 22) argued that, the investigation of the context in Realist Evaluation is very important since programme participants are influenced by the programme context and their acclimatisation to these planned programmes and are influenced by the programme context. Moreover, Realist Evaluation provides a departure from futile ‘one-size-fits-all’ methods of addressing interventions because it is perceptive to differing environmental contexts in programme design and implementation. Realist Evaluation is relevant from formative to summative programme evaluation and through its theory-based foundations, evaluation informs systematic analysis throughout the policy cycle. Stakeholders’ engagement remains integral to the method and they are continuously interrogated for their knowledge of the programme to inform the further refinement of programme theory of change. Whilst such engagements with stakeholders are necessary and welcomed, Realist Evaluation does not assume that stakeholders are infallible experts, and therefore, the validity of all theories is tested against emergent outcome patterns.

Since Realist Evaluation interrogates the programme theory and asks 'what works, for whom, in what context and in what respects', this methodological approach can be applied towards impact evaluation programmes in the public sector that are implemented in large, complex, multi-faceted social environments with little or no understanding of causal mechanism. A programme that is implemented in a different context resulting in different outcomes, even though it was implemented in the same way, could benefit from being evaluated applying this approach.

Weiss (2000: 44) emphatically asserted that theory-based evaluation, which encompasses Realist Evaluation, should not be a regular part of all conducted evaluation, since the method is quite rigorous, elaborate and some evaluations do not require such level of depth and rigour to answer the evaluation questions and may require less probing strategies.

According to Westhorp, Prins, Kusters, Hultink, Guijt and Brouwers (2011: 11-12), Realist Evaluation can be gainfully applied in situations such as where gaining knowledge and insight about the workings of a programme are the aim; where a programme is being implemented in a new context with no previous evidence of how it might work; where a programme is being replicated in another context different from the previous implementation; or in instances where outcome patterns are contradictory from prior implementations. In these instances, the application of Realist Evaluation may serve to ascertain and provide empirical evidence of how the programme works and who can most benefit from it.

3.3.8 The limitations of the Realist Evaluation approach

Realist Evaluation remains as an emergent evaluation method and the application of the approach is continuously evolving. Most matters and concerns about conducting Realist Evaluation in specific settings remain vague; however, the methodological processes and techniques as outlined by Pawson and Tilley (1997a) are applied and provide insights and valuable knowledge on various social policies and programmes (Henry, 2005: 362). This view is supported by other researchers (Marchal et al., 2012) who posited that accurately defining a mechanism and context is open to interpretation and more clarity is required regarding these definitions. In a recent research by Punton, Vogel and Lloyd (2016: 4) there were evident challenges distinguishing between "mechanisms" "features of the context" and "features of the intervention itself". Davis (2005: 291) reinforced this view and claimed that there is poor conceptualisation of what 'context' entails in Realist Evaluation. He stated that this concept of 'context' is understood superficially and is poorly defined and consequently, replication of programmes to other contexts poses challenges.

Moreover, other researchers identified by Marchal et al. (2012) have pointed out that theory-driven evaluations are resource intensive and involve the rigorous testing of programme theory to ascertain its validity on programme context and programme mechanism. Yet others concurred with this view indicating that "If TBE is carried out in full detail, it is apt to be an expensive and time-consuming enterprise" (Birckmayer & Weiss, 2000: 429).

Most recently, Punton, Vogel and Lloyd (2016: 9) found that the theory-driven nature of Realist Evaluation demands flexibility in programme structure and systems, as methods and samples may change in the course of testing and refining programme theory. It was found that this requisite flexibility should not be assumed particularly in the international development sector as systems and processes in those environments may be inflexible.

Others such as Pedersen et al. (2012) argued that the limitation of Realist Evaluation lies in the fact that its evaluation design is not standardised in the same manner as the orthodox RCTs and therein lies some of the challenges.

Evaluators have tended to produce disjointed laundry lists of contexts, mechanisms and outcomes as opposed to configured CMOs that are explanatory. Pawson and Manzano-Santaella (2012:183-189) have identified this practice as flawed and have urged evaluators to produce constructed explanations not catalogues. Pawson and Manzano-Santaella (2012: 184) emphasised that “CMO configuration tests programme theories and to do so the theory must be cast as an *if-then proposition*. A CMO is a hypothesis that the programme works (O), because of the action of some underlying mechanisms (M), which only come into operation in particular contexts (C). *If* the right processes operate in the right conditions, *then* the programme will prevail”.

Along these lines it has also been argued by Astbury (2013: 390) that there are also probable vulnerabilities in forcing adherence to the context-mechanism-outcome prescription with its accompanying tabulation format. Instead of being read in a ‘configured’ or constructed manner the listed contexts, mechanisms and outcomes can be interpreted in a linear manner which may deviate from the explanatory focus of the method. Astbury (2013: 390) then saw a solution to this problem as the initial infusion of the logic model into the realist CMO explanatory theory. On the other hand, Astbury (2013: 389) identified research that found several ‘methodological headaches’ including puzzlement and ambiguity on Realist Evaluation theoretical concepts, challenges in unravelling the linkages amongst programme contexts, mechanisms and outcomes, as well as vagueness regarding when programme theory should be developed and tested. Difficulties involving the handling of complex data, categorising unexpected programme mechanisms and outcomes, managing manifold programme theories, as well as engaging various stakeholders were similarly recognised.

Davis (2005: 292) also reviewed some realist research studies and found that there is a propensity in the applications of the REM to consider only the proposed and specified programme outcomes and disregard programme outcomes that emerge through the process of theory testing resulting in failure of including these emergent outcomes into the evaluation. It is argued then that emergent outcomes after theory testing should be re-incorporated as part of the refined and tested theory to lend credence to the opening of the black box.

Pawson and Tilley (2004) acknowledged that Realist Evaluation has its limitations, since it is intellectually challenging and is far from a tick-box exercise. The approach requires the evaluator to probe programme theory, to outline probable outcomes and to precisely decide what data to research. Once data is found, it is tested and there is arbitration between competing theories.

Quite critically Pawson and Manzano-Santaella (2012: 177) emphasised that if the methodology lacks an explanatory focus on mechanism of change, does not utilise a multi-method data medium, and fails to investigate constructed contexts, mechanisms and outcomes, it certainly is not Realist Evaluation.

Therefore, Realist Evaluation requires advanced theoretical understanding, research design and data analysis skills. Pedersen et al. (2012) agreed and argued that the evaluator's knowledge of the method influences the quality of the evaluation findings which lead to evidence-informed policy decisions.

According to Westhorp et al. (2011: 11-12), Realist Evaluation is ill-advised where gaining insights on programme mechanism and efficacy is not a priority, as the method produces meticulous, and in-depth findings. Similarly, the method demands intensive financial and human resources and in the absence thereof the method should not be adopted. Finally, the envisaged evaluation should be summative in nature, as this will provide a solid traction of data that can be tested empirically to provide evidence on programme context, mechanism and outcome.

3.3.9 Impact Evaluation Assessment Framework

Based on the literature review and the theoretical underpinning of Pawson and Tilley's Realist Evaluation Method of programme evaluation, this research proposes the following Realist Evaluation Impact Evaluation Assessment Framework that provides practical guidelines infusing key elements of the Realist Evaluation approach into practice. This framework may be applied to ensure that impact evaluation implemented in the South African public sector are meaningful, valid and useful to policy-makers. The framework is graphically demonstrated in Figure 3.7 as well as explained in the following sections. This framework will be further refined and a revised model that addresses identified gaps will be presented in Chapter 8.

3.3.9.1 Purpose and objective of the evaluation

The purpose of the evaluation should be clear and coherent and provide a rationale for undertaking the evaluation. This will be derived from clearly defined and specified evaluation questions as well as the 'burning issues' and concerns that the evaluation seek to address. Emanating from this, clear and explicit objectives towards achieving the purpose of the evaluation will be specified. The achievement of the objective should be aligned with the purpose of the evaluation

3.3.9.2 Initial programme theory of change

The programme theory specifies clearly defined pathways that the programme assumes in order to achieve the envisaged change. This theory is 'initial' since it is yet to be confirmed by the emergent programme outcome patterns when implemented. If the theory is well founded, outcome patterns will confirm its resiliency. If the outcome patterns are contradictory to the initial specified theory of change, these emergent insights will be infused into the refined theory, which can be subject to further testing.

As discussed in section 3.3.2 sometimes the programme theory of change is unknown or not explicit enough. Under those condition the programme theory should be reconstructed. It can be reconstructed from interview information sourced from those who designed the programme, the programme practitioners themselves may provide insights on how the programme is supposed to work. If the programme was previously evaluated, such evaluation reports might provide further insights as well as the wider relevant social science literature research. Pawson and Tilley (2004:11) further advise that "documentary analysis, interviews and library searches may all be involved and can help to articulate the formal or official programme theory".

When a programme is planned, its theory of change should explicitly indicate the pathways to change. If these propositions are accurately predicted the observed outcomes should be more or less as envisaged and in harmony with the programme's overarching aims. However, if the programme stakeholders do not respond in accordance with this programme theory, the integrity of the supposed programme implementation chain is weakened.

3.3.9.3 Specified impact evaluation approach and research method

The impact evaluation approach applied in the evaluation should be specified and relevant in achieving the objectives and purpose of the evaluation. The research methods applied can be multi–methods, versatile, pluralist and varied as appropriate, informed by the optimum way of achieving the objectives and purpose of the evaluation. In essence, the research methods applied should serve to assess in-depth interrogation of programme impact.

3.3.10 Utility value of impact evaluation

The utility value of the impact evaluation assesses the ability of the impact evaluation to generate evaluation findings that offer new insights of what works, for whom, under what conditions and in what respects. This is achieved through an explanatory focus that aim to enlighten on what works, for whom, why, how, when and under what circumstances. The explanatory focus is underpinned by specifying how the programme's context, in what is believed to be a complex social system, as well as the programme's mechanism of change contribute to the observed outcomes.

3.3.10.1 The programme context

First, the explanatory focus of the impact evaluation is achieved by establishing the programme context. Contextual conditions under which programmes are implemented are critical. Their surrounding social environments influence social programmes. The same programme will thrive in one social environment and fail in another setting due to surrounding circumstances and other contextual factors. As the success of a programme is strongly dependent on its context, a distinct delineation of the programme context within an assumed complex social system should be specified. The individuals, actors and key agents involved in the programme will enable or impede the implementation of the social programme based on their enthusiasm and will. Another aspect that influence the programme context will be the interpersonal relationship between all programme key stakeholders, such stakeholders will invariably include intended programme beneficiaries, the programme staff, policy-makers and other agents in the implementation chain. The institutional setting surrounding the programme has an influence on programme context. This will be defined by the organisational culture. Furthermore, the infrastructural system is also a key aspect that shape and influence the context of the programme. This entail political backing of the programme, resource allocation afforded to the programme as well the overall public's perception of the programme and due support or opposition to it. Therefore, these contextual factors largely have a bearing on the success of a programme.

3.3.10.2 The programme mechanism of change

Secondly, the programme's mechanism of change is investigated. Mechanism is the "it" factor that result in the occurrence of change. Mechanism is largely influenced by the reasoning and behaviours of the intended stakeholders of the intervention. This change can be engendered in myriad ways, driven by various actors responding in various ways. This mechanism should be defined and tested.

3.3.10.3 The programme outcomes

Thirdly, the observed programme outcomes are assessed. These emergent outcomes are analysed for their relevance to the programme theory. They are scrutinised in terms of whether they confirm the initial programme theory, what can and cannot be explained by the programme theory and whether there are suggestions on further refinement and reframing of the programme theory towards alignment with observed outcome patterns.

3.3.10.4 The CMO configuration

Finally, with this background all three components, the programme context, the mechanism of change and the observed outcome patterns, are constructed together or 'configured' to tell a coherent story line. *The context-mechanism-outcome configuration (CMOC) will specify that under the specified conditions, change will be triggered in a specified manner or mechanism resulting in the generated observed outcomes.* At best, two or three plausible CMOCs should be tested against the specified initial programme theory.

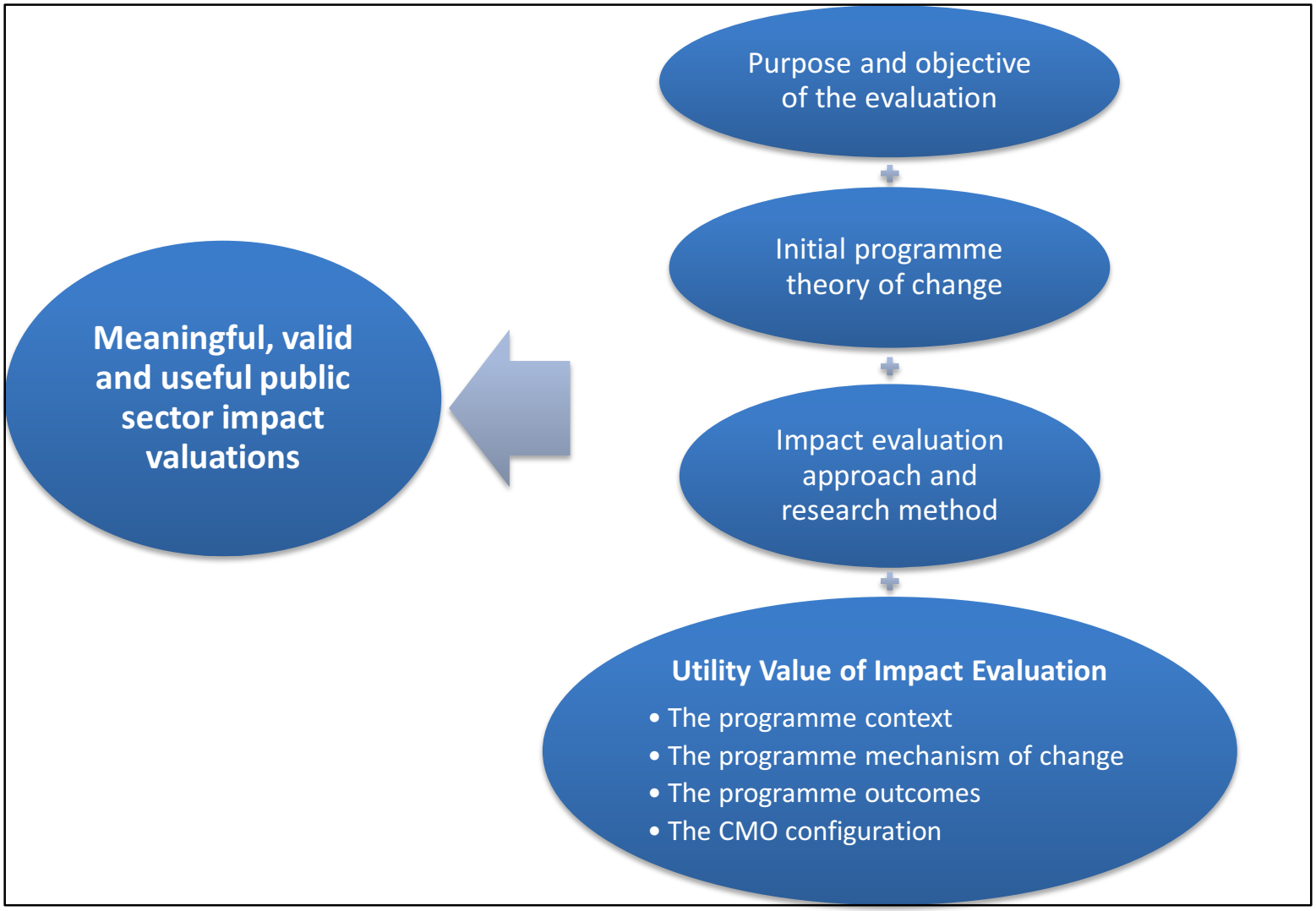


Figure 3.7: Realist Evaluation Impact Evaluation Assessment Framework

Source: Adapted from Pawson and Tilley, 1997.

3.4 SUMMARY AND CONCLUSION TO CHAPTER 3

This chapter provided the current context of the methodological approaches in impact evaluations globally. Stemming from that discussion, the review mapped the development and rationale of Realist Evaluation. In this analysis, the location of the Realist Evaluation Method within the branch of theory-based evaluation was explored. This provided the foundational in-depth understanding of 'theories of change' and how change occurs in programmes, i.e. what Realist Evaluation terms 'programme mechanism'. With this background, the Realist Evaluation methodical approach underpinned by its philosophical approach and its defining key ideas of realist inquiry as well as its research application was a subject of extensive discussion. Realist Evaluation believes all interventions are implemented in complex social environments and a review of the myriad ways and nuances of how social programmes work using the Realist Evaluation lens was therefore explored.

After this, case study examples where Realist Evaluation had been applied in public sector environments internationally were also highlighted. The suitability of Realist Evaluation as method applied in impact evaluations was discussed. The success of Realist Evaluation internationally was reviewed and current trends were reflected upon. This resulted in a synthesis of the limitations of Realist Evaluation from the research evidence.

Following from this discussion, based on the literature review and the theoretical underpinning of Pawson and Tilley's Realist Evaluation Method of programme evaluation, a Realist Evaluation Impact Evaluation Assessment Framework was presented as a theoretical tool of assessing the robustness of impact evaluation that are implemented in the South African public sector in terms of whether such evaluations are meaningful, valid and useful to policy-makers. The next section details the research design methodology as informed by the research objectives.

CHAPTER 4:
RESEARCH DESIGN AND METHODOLOGY

4.1 INTRODUCTION

This chapter outlines the research philosophy and approach. After this, the research design, as informed by the research objectives of the study is presented. Then the research strategy and adopted data collection methods are described including the data analysis, research ethics and limitations of the study.

4.2 THE RESEARCH PHILOSOPHY AND APPROACH

4.2.1 Ontology

Ontology is concerned with what is true and real, therefore the nature of reality. Objective truth is dependent on a philosophy that suggests that our beliefs, whatever they are, have no bearing on the facts or observed phenomena. That which is true remain true irrespective of our belief system or existence (Lee & Lings, 2008:112). Therefore, the researcher's ontological assumptions are an objective stance in the production of valid knowledge.

4.2.2 Epistemology

Epistemology considers the nature of knowledge and different ways of gaining knowledge. This research is positivist in orientation where empirical data is objectively collected and analysed in a structured manner in order to generalise about the phenomena under study. Through the collection of empirical evidence, the emerging facts provide evidence to answer the research question leading to inductive theory building.

Whilst this research is positivist in its orientation, there is acute cognisance that the positivist orientation at its core, espouse a Euro/American worldview which pre-supposes research techniques, data gathering and theorising that promotes the 'academic imperialism' of this worldview. Due to the colonial past of South Africa, the Euro/American evaluation paradigms, conceptual frameworks, standards and approaches are still dominant in the evaluation field, resulting in a one-size-fit-all, monolithic evaluation approaches.

There is therefore an emerging discourse that seeks to challenge this worldview advocating for 'Made in Africa', and 'Africa-Rooted 'evaluation paradigms. This is a decolonisation and indigenisation paradigm advocated by African evaluators such as the panel which convened in Bellagio, Italy to deliberate the issue (Bellagio Report, 2013) as well as theorists such as Chilisa and Malunga (2012), Chilisa, Major & Khudu-Petersen, (2017), calling for African-driven evaluation theories and practices that are decolonised and are essentially indigenous, articulating the ontologies, epistemologies and axiology of the African continent and its people. Evaluation paradigms in this regard, transcend the adaptation or adoption of prevailing methods and

encompass evaluations methods, design, processes and systems and implementation that are authentically and indigenously African.

Such a call is two-fold as it initially necessitates the adaptation of the prevailing methods to be sensitive to the African conditions and contextual factors and through time progressing towards the deliberate and purposeful paradigm shift towards indigenous Africa-rooted evaluation paradigms. Proponents of the adaptation and refinement of the Global North paradigms promotes adaptations since "Africa-rooted evaluation paradigm would not contain substantive differences from the prevailing Western evaluation paradigm, but its purpose, focus, design and implementation would probably just be more sensitive to African cultural contexts and practices in order to achieve the most accurate and valid results" (Cloete, 2016:67). Notwithstanding this state of affairs, there is imploring urgency coming from the Global North (Carden & Alkin, 2012) for a significant engagement by Africans, acting in their context to define contextually relevant methodologies and approaches that crystallise an Africa-rooted evaluation paradigm.

Perceptive and appreciative to the transformative evaluation approaches espoused by the Bellagio Report (2013) and Chilisa and Malunga (2012), and the prevailing decolonisation research agenda, the research whilst positivist in orientation, guided by the research question, adopts methods and designs that are context-sensitive, participative, multi-method encompassing both qualitative and quantitative aspects and integrative in approach.

4.3 RESEARCH DESIGN

The purpose of this research was to find out the potential value of adopting the Realist Evaluation Method approach in programme impact evaluations to ensure that evaluation findings offer insight into what works, for whom, under what conditions and in what respects, thereby enhancing programme impact evaluation practice in the South African public sector.

An adopted Realist Evaluation Impact Evaluation Assessment Framework lens was applied. This theoretical framework, was applied in a basic form to the case studies to make sense of the findings in terms of providing an explanation of why a programme works, in which context a programme may work, how it may work and the envisaged outcomes from such implementation. The analysis framework and its components were relevant in terms of achieving the specific objectives of the study. These objectives are to investigate the methodologies and approaches used in past programme impact evaluations, as well as to establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.

Therefore, the nature of the research is essentially exploratory as new knowledge emanating from the research findings will enrich the intellectual capital of programme impact evaluation in the South African public sector and shed further insight on effective methods of programme impact evaluation. The research question to be answered influences the research design and the strategy adopted.

The research question as presented in Chapter 1 is: **What is the potential value of adopting a Realist Evaluation Method in programme impact evaluations in the South African public sector? Can such an approach offer new insights of what works, for whom, under what conditions and in what respects, thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making?**

The specific objectives are:

- To gain an in-depth understanding of the Realist Evaluation Method through a detailed review and analysis of the related literature.
- To assess the current trends in research and application of Realist Evaluation methodical approach in conducting impact evaluations.
- To investigate the methodologies and approaches used in past programme impact evaluations in the South African public sector.
- To establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.
- To establish the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector.

The overarching research design detailed here is also graphically presented in Figure 4.1 below.

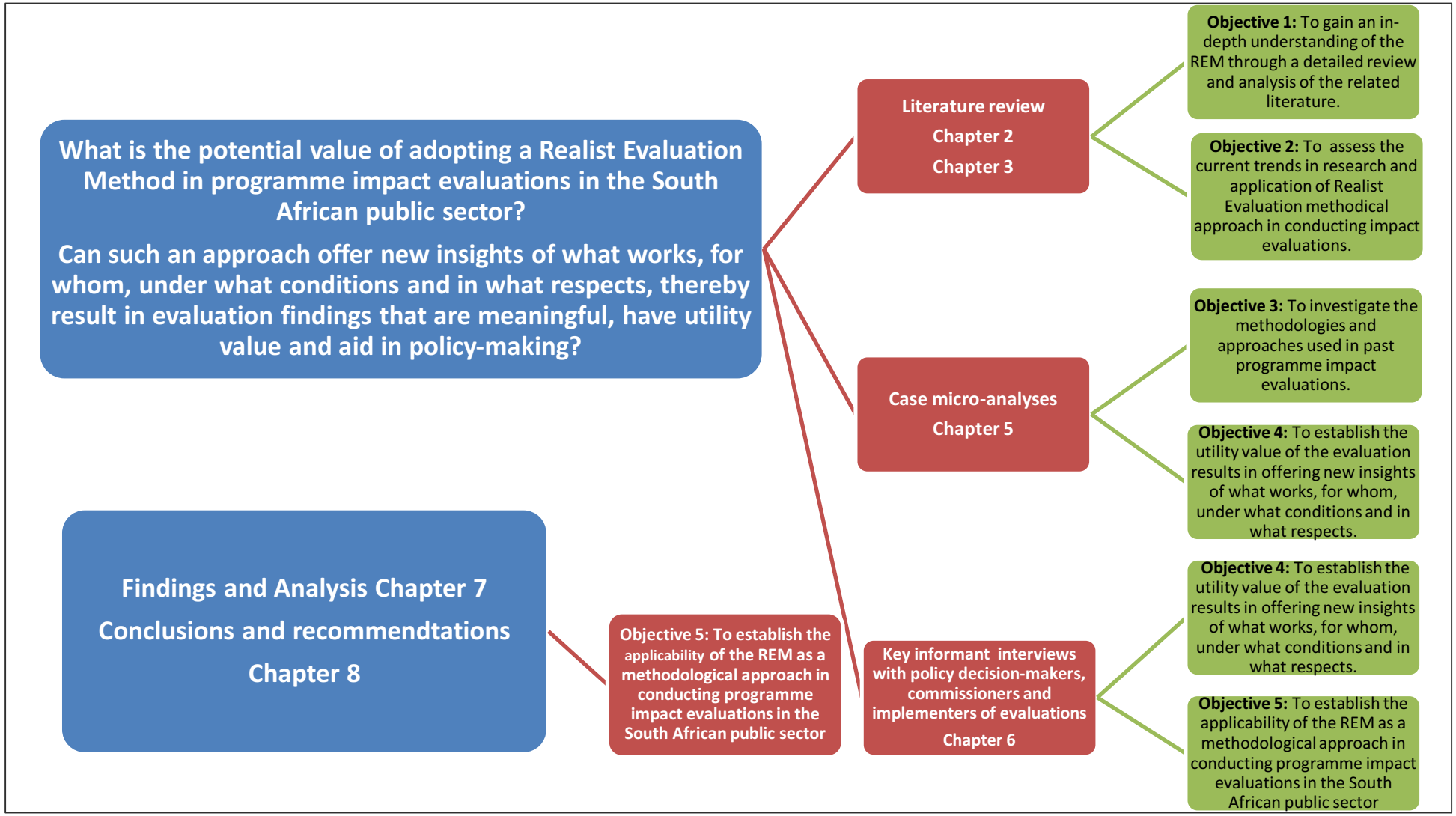


Figure 4.1: The research design

Source: Author.

4.4 RESEARCH STRATEGY AND METHODS

This section describes the strategy adopted to undertake the research and the data collection methods used to investigate the research objectives. The precise balance of methods used was selected in accordance with the research question tested and with the available data. The research had a three-pronged approach inclusive of comprehensive literature review, case study micro-analyses of past programme impact evaluations within the South African public sector and key informant interviews with policy-makers, commissioners and implementers of evaluations in South Africa

4.4.1 Literature review

A comprehensive literature review that provided a strong theoretical base was conducted with the aim of gaining understanding and insights on trends in impact evaluation including theory and practice globally and locally as well as the specific evaluation practice in the South African public sector. The emergence and international practice of the Realist Evaluation Method within the context of programme evaluation globally was also presented.

4.4.2 Analysis of literature review

A critical analysis of available literature was conducted and presented in Chapters 2 and 3 in order to pursue the related research aims and objectives. Both recent and seminal works on evaluation and Realist Evaluation were explored in order to present key and relevant aspect of the literature as pertaining to the research objectives and research question.

4.4.3 Case study selection

The aim of conducting case study analysis is to establish the methodologies and approaches used in past programme impact evaluations within the South African public sector. Moreover, research (Rogers & Peersman, 2014: 85) has called and argued for an impact evaluation research agenda that focuses particular attention to detailed, theory-informed, mixed-method comparative case studies of the actual processes and impacts of impact evaluation. Others have highlighted that case study analysis is well suited in verifying the Realist Evaluation prescript of ‘what works, for whom, in what context and in what respects’ (Woolcock, 2013: 16).

According to Yin (2009: 260), multiple cases covering different contextual conditions significantly expand the generalisability of findings to a broader collection of contexts compared to a single-case study. Largely, the evidence from multiple case studies produces a more compelling and robust case study.

To meet this objective, case studies of programme impact evaluations conducted in the South African public sector within the period of the year 2000 until 2015 were searched, identified and selected. The rationale for selecting the year 2000 as a starting point is based on the premise that the Realist Evaluation Method as an approach in conducting programme evaluation emerged in the

late 1990s and solidified in nearly two decades through various works such as those of Pawson and Tilley (1997a) and Pawson (2006; 2013). Therefore, since the REM only emerged in 1997, it will not be relevant to look for studies prior to 2000.

The search for impact evaluations commissioned by public sector institutions was conducted in the DPME portal of all completed evaluations within the National Evaluation System (NES). A large number of evaluation from before the establishment of the NES are in the portal. These pre-NES evaluations have gone through quality assessment for inclusion within the portal. The other search criteria included key terms such as ‘impact evaluation’ and ‘impact assessment’. Cases in the portal which did not meet this criterion were excluded.

The search elicited one pre-NES impact evaluation and two post-NES impact evaluations that was undertaken in the South African public sector. A total of three cases were found in the DPME portal that met the search criteria were selected for inclusion in this study. The search results were verified via e-mail with the DPME’s evaluation unit to ensure that further impact studies were not available, but perhaps not placed on the portal. It was confirmed that the found three cases were the only impact evaluations completed in the NES, although two further studies had been commissioned by the DPME. A DPME official confirmed this by email correspondence on 6 May 2016 (Masikane, 2016). Of the of three case studies available on the repository, one was still in the process of being completed and the final evaluation report was not yet approved by Cabinet by the time this research process was completed. It was therefore withdrawn from the case study list. Similarly, at the time of this research, the further impact studies commissioned under the NES were not completed and/or approved by Cabinet and these studies were also excluded from this study.

It was further premised that there could be other impact evaluations that have been commissioned by the South African public sector outside the DPME repository. Having exhausted the further availability of impact evaluations conducted by the state within the DPME evaluation repository an international search was conducted for evaluations conducted by international development institutions. Working from a list of international development institutions, a search was done on the websites of international institutions such as UNICEF, the World Bank, International Monetary Fund, and the International Initiative for Impact Evaluation (3ie) to identify possible impact evaluations completed in South Africa. This was effectively an international search for state commissioned completed impact evaluations conducted by international development institutions. Search terms such as ‘South Africa’, ‘impact evaluation’ and ‘impact assessment’ were used. This search elicited an impact evaluation conducted by the World Bank as well as another one conducted by the International Initiative for Impact Evaluation (3ie) for South African public sector programmes. Further searches did not elicit any additional completed evaluations.

The NES repository and international donor website reviews effectively presented in a total of four state commissioned impact evaluations, two from the DPME repository and one from the World Bank and another one from the International Initiative for Impact Evaluation (3ie). Given this limited number

of impact case studies publicly available, all four cases identified were included in the research (see Chapters 5 and 7).

4.4.4 Case study analysis

The process for case study analysis is an adapted Realist Evaluation Impact Evaluation Assessment Framework. The framework is adapted from Realist Evaluation theoretical framework of Pawson and Tilley (1997a) and Pawson (2006; 2013), as well as the literature review. For each case study, the framework will be applied as a lens to present the data collected in each case in terms of the programme context, mechanism and outcomes. The Realist Evaluation Impact Evaluation Assessment Framework and its components are fully described in Chapter 5. This theoretical framework enabled data analysis and interpretation that led to answering the question of what would be the value of a Realist Evaluation approach to produce findings that offer new insights of what works, for whom, under what conditions and in what respects, and thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making.

4.4.5 Key informant interviews

4.4.5.1 Overview

Several key informant interviews were conducted with respondents who are familiar with the impact evaluations selected as case studies and who can reflect on the specific information needs of policy-makers as a user of the evaluation reports.

The first objective of the key informant interviews was to establish from policy decision-makers, commissioners and implementers of evaluations in the South Africa public sector the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects and thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making.

The second objective of the key informant interviews was to establish from policy decision-makers, commissioners and implementers of evaluations in the South Africa public sector the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector.

4.4.5.2 The selected respondents

The selection of respondents was informed by the identified and available impact case studies described in preceding section. Respondents were purposively selected based on their presumed knowledge and expertise of the selected impact case studies or sectors included in this research. The selected respondents and key informants were commissioners of the cases under this study. Most of whom were involved in the design of policy evaluations in basic education, social protection and human settlements. These were experts who could provide policy insights in these policy areas.

Others selected were policy drafters and policy advisors who could provide key policy insights in basic education, social protection and human settlements. In addition, the sector specialist, known as Outcome Facilitators, in basic education, social protection and human settlements were sought as one of the key respondents in this study. These respondents monitor the implementation of policy outcomes, in this study the key ones being the quality of basic education, an inclusive and responsive social protection system as well as the implementation and coordination of sustainable human settlements and improved quality of household life.

Five respondents that could offer perspectives on policy interventions in basic education were selected. Additional six respondents that could offer perspective and insights on policy interventions in social protection were selected. Finally, four key experts that could that could offer perspective and insights on policy interventions in human settlements were selected. This resulted in a total of 15 respondents who were approached for access and cooperation. Table 4.1 below provides a summary of the respondent groups per case study.

The final seven respondents were as follows: two offered perspectives on policy interventions in basic education, three offered perspectives and insights on policy interventions in social protection, one sector specialist offered insights on human settlements and one expert offered insights on economic cluster evaluations. Representation in the policy area of basic education was balanced as the views of both the commissioner of education evaluations as well as the sector expert were valuable in providing perspectives on commissioning and implementation of evaluations as well as key education policy insights and perspectives. On the other hand, the policy area of social protection in terms of the CSG was somewhat over-represented as there were two commissioners who had oversight over the same evaluation, the CSG. However, their valuable views were well balanced by the inputs of an external policy expert who provided broader perspectives and valuable insights especially on ECD, the CSG and Grade R evaluation.

The policy area of human settlement was underrepresented due to the unavailability of the respondents. In particular, gaining access to commissioners in the human settlements government department was unsuccessful. However, this gap was closed as the human settlement sector expert, from the DPME, effectively provided valuable insights on government human settlement evaluations, the implementation of human settlement policy outcomes, as well as policy priorities in the implementation and coordination of sustainable human settlements. In this regard, a single interview was deemed adequate as the breadth and scope of information provided was relevant and valid in furthering the research objectives. In addition, the interview information was triangulated with a search for specific parliamentary questions, which were asked by policy-makers in relation to the selected human settlement impact evaluation. The relevant parliamentary questions and answers provided are included in Appendix J.

Another underrepresented area was the economic cluster evaluation, the youth wage subsidy. This required an expert linked to the National Treasury who had presumed knowledge and expertise on

labour market interventions. While only one respondent participated in this part of the evaluation, the views of this respondent are regarded as highly credible given his involvement in the design of the specific youth wage policy evaluation and part of the international team commissioned by the National Treasury. The views and perspectives of the expert were most valuable in understanding the circumstances surrounding the implementation of the youth wage policy evaluation as well as the key policy decisions and priorities of government regarding youth unemployment policies. In this regard, a single interview was deemed adequate as the breadth and scope of information provided was relevant and valid in furthering the research objectives. In addition, the interview information was triangulated with a search for specific parliamentary questions, which were asked by policy-makers in relation to the youth wage subsidy impact evaluation. The relevant parliamentary questions and answers provided are included in Appendix G and National Treasury statement on the current status of the youth wage subsidy on Appendix H. The researcher is of the view that the pool of final respondents provided valuable inputs which helped to advance the study in answering the research objectives.

Table 4.1: Summary of respondent groups

No	Respondent	Designation	Department	Policy Area	Case Study	Interview secured?
1	Commissioner Education Evaluations	Director: Early Childhood Development	Department of Basic Education, South Africa	Policy interventions in basic education	*The impact of the introduction of Grade R on learning outcomes*	No
2	Sector Expert Education Evaluations	Advisor	Department of Basic Education, South Africa	Policy interventions in basic education	*The impact of the introduction of Grade R on learning outcomes*	Yes
3	Commissioner Education Evaluations	Chief Director-Planning, Research and Coordination	Department of Basic Education, South Africa	Policy interventions in basic education	*The impact of the introduction of Grade R on learning outcomes*	Yes
4	Sector Expert Education Evaluations	Outcomes Facilitator, Outcomes 1 and 5	Department of Planning, Monitoring and Evaluation, South Africa	Policy interventions in basic education	*The impact of the introduction of Grade R on learning outcomes*	No
5	Education Policy Expert	Professor of education policy and advisor to the Gauteng education department	University of Witwatersrand School of Education	Policy interventions in basic education	*The impact of the introduction of Grade R on learning outcomes*	No
6	Policy Expert Social Development Evaluations	Former Director: Children: National Department of Social Development	Child Welfare South Africa	Policy intervention in social protection	*The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households*	Yes

No	Respondent	Designation	Department	Policy Area	Case Study	Interview secured?
7	Policy Expert Social Development Evaluations,	Advisor Minister: Social Development	Department of Social Development	Policy intervention in social protection	*The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households*	No
8	Commissioner Social Development Evaluations	Director: Evaluations	Department of Social Development, South Africa	Policy intervention in social protection	*The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households*	Yes

Table 4.1: Summary of respondent groups (continued)

No	Respondent	Designation	Department	Policy Area	Case Study	Interview secured?
9	Commissioner Social Development Evaluations	Head: Monitoring and Evaluations	Department of Social Development, South Africa	Policy intervention in social protection	*The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households*	Yes
10	Policy Expert Social Development Evaluations	Outcomes Facilitator, Social Protection	Department of Planning, Monitoring and Evaluation, South Africa	Policy intervention in social protection	*The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households*	No
11	Policy Expert Human Settlements Evaluations	DDG: Human Settlements Planning and Strategy	Department of Human Settlements, South Africa	Policy interventions in human settlements	*An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa*	No
12	Commissioner Human Settlements Evaluations	Chief Director, Monitoring & Evaluation	Department of Human Settlements, South Africa	Policy interventions in human settlements	*An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa*	No
13	Commissioner Human Settlement Evaluations	Director, Impact Assessment	Department of Human Settlements, South Africa	Policy interventions in human settlements	*An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa*	No
14	Sector Expert Human Settlements Evaluations	Outcomes Facilitator, Outcomes 8, (Human Settlements)	Department of Planning, Monitoring and Evaluation, South Africa	Policy interventions in human settlements	*An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa*	Yes
15	Policy Expert	Economic Cluster	Stellenbosch University	Policy intervention in social protection	*Youth Wage Subsidy experiment for South Africa*	Yes

Overall, these respondents were deemed to bring relevant expertise, insights and perspectives on whether adopting a Realist Evaluation approach in programme impact evaluation in the South African public sector would result in evaluation findings that offer new insights of what works, for whom, under what conditions and in what respects, and thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making.

Out of the sample of 15 respondents who were approached, seven responded and granted access through face-to-face interviews. This resulted in a response rate of 46.6 percent. Two respondents were interviewed on policy perspectives in basic education. Another four respondents were interviewed on policy insights in social protection. A final one was interviewed on policy insights in human settlements. Although these respondents were selected on the basis of their familiarity with a specific impact evaluation case study, the nature of the questions allowed respondents to offer inputs relevant to all four cases included in this research study.

Gaining access to key respondents such as the relevant ministerial advisor who would have policy insights on some of the policies informing the impact evaluations in this study as well as some of the DPME outcome facilitators who monitor the implementation of policy outcomes was difficult. Through persistence following up, in the final stage there were a total of seven respondents who granted access and were interviewed. Whilst this is a limited number, these were policy experts in the relevant areas of this study who commissioned evaluations and drafted policies. Three were experts in basic education policy evaluations, two were experts in social development and had commissioned the relevant evaluation that is part of this study, one was a DPME government outcome facilitator who provided broad perspective and insights on the role of impact of evaluations in government in informing progress on policy outcomes as well as human settlements policy evaluations. The last respondent was an expert who was involved in youth employment policy evaluations who provide key insights on the relevant evaluation that is part of this study as well as current perspectives on the implementation of youth employment policies.

Whilst this limited number of interviews possibly provided fewer perspectives and insights, these respondents were able to offer input covering all four case studies. Some of the gaps in policy insights were mitigated by a search for specific parliamentary questions, which were asked by policy-makers in relation to the selected impact evaluations in this study. The relevant parliamentary questions and answers provided are included in Appendices F, G, I and J.

4.4.5.3 The interview research questionnaire

A questionnaire was developed as an additional data collection technique to further substantiate and triangulate the findings from the literature review and the case study analyses.

The research objectives guiding the interviews were to:

- To establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.

- To establish the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector.

Therefore, the research objectives informed the identification of the data requirement and the design of the questionnaire through drawing up of a list of appropriate questions. Therefore, key data requirements informing these objectives was to find out:

- What do policy-makers want?
- Do they value evidence?
- What questions are they asking?
- What makes evaluation findings meaningful, valid and useful, from their perspective?
- What are the most important limitations with existing policy impact evaluations?
- Are evaluation approaches adopted in general in the public sector always appropriate to inform their needs?
- Is there a potential value in the application of the Realist Evaluation approach?
- What are the potential drawbacks or negative implications of adopting a Realist Evaluation approach?

Questionnaires that produce reliable and valid information according to Hair Jr, Celsi, Money, Samouel and Page (2011: 249) prioritise the careful consideration for designing the questionnaire, the clarification of key concepts that remove ambiguity, determining the type of questions to ask and their sequence, pretesting the questionnaire and final administering of the questionnaire. These steps were carefully considered toward the final utilisation of the questionnaire. Reference is made here to some of the questions that were included in the questionnaire to capture the various aspects, whilst the research questionnaire is included as Appendix E.

It is acknowledged that the key informants were not the direct policy-makers as they do not have policy-making powers. However, they have the authority to influence the direction of policy-making as well as implement the relevant policies. Therefore, the sample of key informants contribute to the policy-making life cycle. The selected sample of respondents were the closest proxy to the actual policy-maker. In this regard the questionnaire catered and guarded against overall bias through the appropriate framing of the questions. For example, Question 2 in the questionnaire asked “In your opinion, what would policy-makers deem to be the 5 most meaningful, valid and useful important aspects in this list?”. Next, Question 3 further asked “Why do you think policy-makers would regard these aspects to be most important?”. Also, Question 4, asked “I would imagine that these expectations may vary between policy-makers and evaluations. *Do you think* all policy-makers would agree with your ranking above? Therefore, the questions posed enabled the respondent to reflect on the overall ‘state of mind’ and expectations of the policy-maker.

It is possible that some of the respondents could have been biased given that some were directly involved in the commissioning of the case study evaluations. Therefore, they could have a vested interest in how the evaluation outcomes are considered. However, these views were substantiated and balanced with the views of external experts not directly involved with the evaluations as well as policy-maker views through the parliamentary questions and answers. Therefore, the research can draw very limited conclusions from the interview findings due to the limitations of the interview sample however these are relevant and adequate for the aims of meeting the objectives of the study.

4.4.5.4 The negotiated access

A letter of introduction was sent to each identified respondent. The letter introduced the researcher, the objectives of the research, the value of the research to the work of the organisation, request for a face-to-face interview with the researcher or the completion of the attached questionnaire. A sample letter is attached as Appendix D in this regard.

4.4.5.5 The key informant interview

Key informant meetings that included prepared questions were held with the identified respondents. The researcher took notes during the interviews.

4.4.6 Interview analysis

Whilst the first few questions from the interview questionnaire were quantitative, the analysis of the data from the key informant interviews was mostly qualitative in nature. This required the identification and analysis of emerging patterns and themes in the data in order to ascertain how these support the research objective. The quantitative analysis was analysed with Microsoft Excel including graphs and tables. Each questionnaire was coded and transcribed into an online survey software tool, SurveyGizmo, which was used to analyse the qualitative data. The tool thematically analysed the data, identifying similar constructs and phrases and generated a standard report.

4.5 RESEARCH ETHICS

The researcher adhered to the Stellenbosch University research ethics and protocols in conducting this research. The Stellenbosch University process on ethical research included familiarity with ethical codes of conduct as detailed in the Policy for Responsible Research Conduct at Stellenbosch University. Steps to ensure established ethical standards were applied including written appropriate informed consent. Participants were also informed that they have the right to refuse to answer questions and they have the right to withdraw from participation at any time. Steps were also taken to ensure personal data of respondents is secured from improper access and that confidentiality of data is maintained. The research met all criteria of trustworthiness, in terms of credibility, dependability and confirmability.

4.6 SUMMARY AND CONCLUSION TO CHAPTER 4

This chapter presented the overarching structure of the research design and adopted methodology. It was highlighted that the nature of the research is essentially exploratory, as new knowledge, emanating from the research findings will enrich the intellectual capital of programme impact evaluation in the South African public sector and shed further insight on effective methods of programme impact evaluation. The research had a three-pronged approach inclusive of comprehensive literature review, case study analyses and key informant interviews with policy-makers, commissioners and implementers of evaluations in the South African public sector.

The next chapter focuses on micro-analyses of the selected case studies focusing on policy interventions in social protection, basic education and social housing.

CHAPTER 5:
POLICY INTERVENTIONS IN SOCIAL PROTECTION, BASIC EDUCATION AND SOCIAL HOUSING

5.1 INTRODUCTION

The focus of the chapter is presenting the case data. As such, for each case the adopted Realist Evaluation Impact Evaluation Assessment Framework lens, fully described in Chapter 3, section 3.3.9 and graphically presented in Figure 3.7, is applied. The theoretical framework searches for the purpose and objective of the evaluation, the articulation of the programme theory of change as well as the specified evaluation approach and research method. In addition, the utility value of the evaluation is assessed by first establishing the programme context. Secondly, the programme's mechanism of change is investigated to establish whether this is defined and tested. Thirdly, the observed programme outcomes are assessed on their relevance to the programme theory. These components, the programme context, the mechanism of change and the observed outcome patterns offer an explanatory focus through a constructed or 'configured' coherent story line. *The context-mechanism-outcome configuration (CMOC) will specify that under the specified conditions, change will be triggered in a specified manner or mechanism resulting in the generated observed outcomes.*

The theoretical framework is used to present the case data and make sense of the findings in terms of providing an explanation of why a programme works, in which context a programme may work, how it may work and the envisaged outcomes from such implementation. The analysis framework and its components are relevant in terms of achieving the specific objectives of the study which are to investigate the methodologies and approaches used in past programme impact evaluations as well as to establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.

This chapter presents the micro-analyses of four case studies. The first two cases are policy interventions on social protection namely, *The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households* (RSA, 2012) and the *Youth wage subsidy experiment for South Africa* (Levinhson, Rankin, Roberts & Schöer, 2014). These are followed by a case study on policy intervention in basic education, *The impact of the introduction of Grade R on learning outcomes* (Van der Berg, Girdwood, Shepherd, van Wyk, Kruger, Viljoen, Ezeobi & Ntaka, 2013) as well as a case study on a social housing policy intervention, i.e. *An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa* (RSA, 2011b).

5.2 POLICY INTERVENTIONS IN SOCIAL PROTECTION

5.2.1 Case study 1: The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households (RSA, 2012)

The case presentation as detailed below is drawn from the above case document as well as the qualitative study report, *Child Support Grant Evaluation 2010: Qualitative Research Report* (RSA, 2011c). All references where indicated are drawn from these documents.

5.2.1.1 Case background

The Child Support Grant (CSG) is a critical element of social welfare in South Africa, benefiting over 10.7 million children monthly as at 2012. The CSG was first introduced in 1998. Between then and 2012, the social grant programme evolved into one of the most comprehensive social protection systems in a developing country. The aim of the support grant is to target children most in need and those in vulnerable and poor conditions. After much policy advice, the age eligibility criterion and the means test for qualification were relaxed and this resulted in increased uptake, greater traction and perceived improved equity in accessing the Child Support Grants (RSA, 2012: 1-2).

The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households of May 2012 (RSA, 2012) was analysed for the purposes of this research. Prior to this impact assessment as qualitative component, which served as preparatory work and a "design evaluation" was conducted in 2011 titled *Child Support Grant Evaluation 2010: Qualitative Research Report* (RSA, 2011c) was performed to inform the preparation of this impact assessment. The case was commissioned by the National Department of Social Development, the South African Social Security Agency (SASSA) and conducted by the department and the United Nations Children's Fund (UNICEF).

5.2.1.2 Scope of the evaluation

The geographic scope of the CSG impact assessment covered households in five provinces, i.e. the Eastern Cape, Gauteng, KwaZulu-Natal, Limpopo and the Western Cape, evaluating children and adolescent beneficiaries receiving and not receiving the CSG (RSA, 2012: ii).

5.2.1.3 Objectives of the evaluation

The rationale for the evaluation was to provide evidence of the impact of the CSG on its beneficiaries, the children. Previous research (Agüero, Carter & Woolard, September 2007; Budlender & Woolard, 2006; Delany, Ismail, Graham & Ramkisson, June, 2008; Makiwane & Udjo, 2006; Samson et al., 2008 in RSA, 2012: 1) highlighted the impact of the CSG in reducing poverty and promoting better human development outcomes through better nutrition, improved school attendance and health. However, previous evidence was at household level and new evidence with rigour directly from the children beneficiaries were required by government (RSA, 2012: 1).

The objectives of the evaluation as defined in the evaluation report were defined by the following research questions (RSA, 2012: 3):

Question 1: How has early versus late enrolment affected the well-being and development of children? In particular, the study addresses this question in terms of children's anthropometry, health and schooling, as well as their access and use of preventative health and nutrition care.

Question 2: How are critical life course events of adolescents affected by the extension of the CSG? Specifically, this study explores this question in terms of adolescents' participation in risky behaviours, schooling outcomes, and work inside and outside the home.

Question 3: What conditions determine and influence access to the CSG? This study focuses on this question at the point of initial application, by assessing the duration and continuity of receipt at the same time as analysing current access and use.

5.2.1.4 Hypothesis and theory of change

The CSG evaluation has a defined programme theory of change mapping the programme's pathways to change (RSA, 2012: 6-7). The hypothesis analysed from documentary analysis sought to test and confirm that (RSA, 2012: 3-5):

- Cash grants targeted at children directly reduce the poverty and vulnerability of children living in poor households;
- The CSG both increases consumption and enables poor households and carers to participate in productive economic activity (e.g. to look for work);
- The CSG addresses the underlying causes of poverty, by enabling poor households to invest in physical, social and human capital assets (education, health, nutrition), that can generate future streams of income; and
- Receipt of the CSG reduces risky behaviour by adolescents, such as transactional sex, alcohol consumption and substance abuse;
- Specific features of the CSG (including that it is unconditional, that it targets caregivers, that it is delivered periodically and predictably, and that transaction costs are relatively low) all ensure that the overall net effectiveness of the programme's social and economic impact is maximised.

5.2.1.5 Impact evaluation methodology

The study employed a mixed-method quasi-experimental design, matching methods to establish attribution of impacts to early versus late enrolment. According to cases analysis "the evaluation employs non-experimental approaches rather than a randomised experiment because there is no practical or legal scope for randomly allocating grants in South Africa (RSA, 2012: ii). Participants in the study were separated into 'treatment' and 'control' groups based on whether or not they were receiving the grant and the duration of receipt. Participants from each group were matched using a

propensity score, which in turn was based on observable characteristics. The comparison of participants from each matched pair resulted in the estimation of dosage-response to a specific number of years' exposure to the grant, in terms of important outcomes such as schooling.

Participants in the study were separated into statistical 'treatment' and 'control' groups made of households with similar observable characteristics that influence their probability of application for receipt of the CSG, based on whether or not they were receiving the CSG and for what length of time they had been receiving the grant. Members from each group were matched using a propensity score, which in turn was based on observable characteristics. By comparing the members of each matched pair, researchers were able to estimate what the dosage-response to a specific number of years' exposure to the CSG was in terms of important outcomes such as schooling.

5.2.1.6 Data collection

Surveyed households and enumerators filled out three questionnaires. One focused on the entire household, while the other two focused in depth on the sampled young child or adolescent. In addition, adolescents completed a confidential, self-administered survey (RSA, 2012: 13). The household questionnaire collected detailed information on household characteristics at the time of the child's birth as well as details about their current living situation and context. These included measures of wealth, household demographic structure, characteristics of the caregiver, location characteristics and access to forms and offices needed to apply for the CSG. This data provided covariates that could be used to match early and late enrolees.

The second questionnaire contained a detailed set of questions on when the household enrolled in the CSG and whether access to the CSG was interrupted. It also collected detailed information on schooling histories, child time allocation and anthropometry. The children also completed tests that covered reading and mathematics skills. The final adolescent questionnaire was designed and administered to a sample of 15-, 16-, and 17-year-olds and their households. The adolescent questionnaire likewise included questions on demographics, schooling history, labour, time allocation, access to CSG and other questions that permitted comparisons of beneficiaries and non-beneficiaries and of youth just above and below the age eligibility cut-off to identify CSG impacts. The confidential, self-administered survey completed by adolescents included questions about their receipt of the CSG, school and work participation, and their engagement in risky behaviours.

5.2.1.7 Data collected on context

Analysis of case documents detailed information on household characteristics at the time of the child's birth, as well as details about their current living situation and context. Contextual factors included measures of wealth, such as saving behaviour, access to banking facilities, household demographic structure, characteristics of the caregiver, location characteristics and access to forms and offices needed to apply for the CSG (RSA, 2012: 13). Actors and agents in the implementation

chain were also interviewed. These included officials from SASSA, health workers, education workers, community leaders and the caregivers (RSA, 2011c: 16-17).

5.2.1.8 Data collected on mechanisms

The case sought to find out whether the intervention works or not; in this instance, the impact of the CSG on the early life circumstances of beneficiaries. The differences in the life circumstances of those who received the CSG and those who did not were studied. Mechanisms refer to the ways in which change occurs. The underlying reasons leading to achieved outcomes were not explicit and therefore the mechanisms of the intervention in terms of how and why they worked were missing.

5.2.1.9 Data collected on outcome patterns

The outcomes anticipated as informed by the initial hypotheses are that CSGs result in reduction in poverty and vulnerability of children beneficiaries, living in poor households. To this end, information on indicators of wealth, such as access to electricity, recipients living in homes with corrugated roofs, access to banking, was collected (RSA 2012: 16). Other outcome studies were the effect of the CSG on household economic consumption patterns and behaviour (RSA, 2011c: 11) as well as the behaviours of the adolescents who received the CSG (RSA, 2011c: 79).

Having presented the case study in this section, its assessment will be done in the following Chapter 6.

5.2.2 Case study 2: A youth wage subsidy experiment for South Africa (Levinhson et al., 2014)

The case presentation as detailed below is drawn from the above case document and all reference are drawn from the same case document.

5.2.2.1 Case background

Due to the high unemployment, rate amongst African youth, which was estimated at above 60 percent in 2012 in the 20- to 24-year age group, the South African government, explored interventions that could incentivise employers to hire young people. Two policy options were proposed, namely, an employment tax incentive where employers can claim tax credits for youth employed for at least two years, or a hiring voucher where youth will present vouchers to hiring employers who can then claim them (Levinhson et al., 2014: 10-11).

Subsequently the tax incentive option was adopted by government and implemented in January 2014. This case study focuses on the policy option of an employer youth wage subsidy, where youth will present vouchers to hiring employers who can then claim the subsidy (Levinhson et al., 2014). The study was commissioned by the South African National Treasury and the National Department of Labour and undertaken by the International Initiative for Impact Evaluation (3ie).

5.2.2.2 Scope of the evaluation

According to the case study "project was implemented prior to the National Treasury releasing details of how the youth employment tax incentive would be implemented nationally. The Employment Tax Incentive policy implemented on 1 January 2014 is a tax incentive for the hire of young people. It lasts for up to two years and is a greater amount than the amount in this project." (Levinhson et al., 2014: 11). The young people in the study were African males between the ages of 20 and 24 years old, had a Grade 12 or matric high school education, and were drawn primarily from the Gauteng province followed by a proportion in the Limpopo province and KwaZulu-Natal province. The potential employers were private companies that were registered, tax compliant and made unemployment insurance contributions for their employees (Levinhson et al., 2014: 54-55).

5.2.2.3 Objectives of the evaluation

According to the documentary analysis (Levinhson et al., 2014: 11), the questions the evaluation sought to answer were the following:

- Are those with a wage subsidy more likely to be in employment as a result of the allocation of the voucher?
- If yes, what are the mechanisms through which this effect works?
- Do voucher holders have different types of jobs compared to those of non-voucher holders?
- Does a voucher's effect persist after it has lapsed? Are there any discernible differences in the employment probabilities between voucher holders and non-voucher holders two years after voucher allocation?
- How do firms respond to the voucher? Can their reactions inform the debate around the implementation of the wage subsidy?

5.2.2.4 Hypothesis and theory of change

According to the case study (Levinhson et al., 2014: 13) the theory of change of the intervention is that:

- A school-leaver is allocated a voucher that enables any firm (subject to the firm being registered for tax and paying unemployment insurance) that decides to employ this worker to claim back a portion of the wage that the firm pays to the worker.
- This young person searches for a job through the channels that are available to them, including their networks, formal application procedures, and informal methods such as approaching firms directly.
- The firm chooses to experiment with an additional worker who is unable to signal their productivity, knowing that the cost of employing this worker is reduced by the amount of the subsidy (less the administrative cost of claiming the subsidy). Through this employment, the worker gains skills and references that increase their productivity and ability to signal this productivity, which raises their income and the likelihood of being retained in employment.

- The firm not only increases the productivity of its workforce, but also raises the productivity and reduces the uncertainty associated with the available pool of young workers.

5.2.2.5 Impact evaluation methodology

The study employed a RCT matching methodology where a randomly-selected sample of 4 009 matched pairs young people were randomly allocated to either the treatment or control group based on specified criteria. They were given wage subsidy vouchers and had the process of claiming these vouchers explained to them during 2010. Both the treatment and control groups were subsequently interviewed in 2011 and 2012 and their performance in the labour market was documented.

5.2.2.6 Data collection

Data was collected through administered structured personal interviews with sampled individuals, focusing on demographic and household characteristics, educational qualification, and previous and current labour market experiences (Levinhson et al., 2014: 15).

5.2.2.7 Data collected on context

Contextual factors focused on individual demographics, their household characteristics, educational qualification, and previous and current labour market experiences of the young people (Levinhson et al., 2014: 15). The young people were African males between the ages of 20 and 24 years, who had a Grade 12 or matric high school education, and were drawn primarily from the Gauteng province followed by a proportion in the Limpopo province and KwaZulu-Natal province. The potential employers were private companies that were registered, tax compliant and made unemployment insurance contributions for their employees.

5.2.2.8 Data collected on mechanisms

The case sought to test whether the intervention works or not; in this instance, the impact of a youth wage subsidy in the form of a voucher held by unemployed youth who give the voucher to employers who then claim the subsidy if they offer the individual wage employment.

The differences in the labour market outcomes of those who received the wage subsidy in the form of a voucher and those who did not were studied. Mechanisms refer to the ways in which change occurred, how it occurred and why. In this case, these underlying reasons leading to achieved outcomes were not explicit and therefore the mechanisms of the intervention in terms of how and why it worked were missing and as such, there is no evidence of collected data on appropriate mechanisms.

5.2.2.9 Data collected on outcome patterns

The outcomes anticipated, as informed by the initial hypotheses, are that those with a wage subsidy are likely to be in wage employment because of the allocation of the voucher. Data was collected on actual use of the wage subsidy voucher by the school leaver by measuring the labour participation

rate of the voucher holders. Data was also collected on wage employment tenure and earnings because of the voucher. Data was also collected on gender and wage employment as well as matric and wage employment.

Having presented the case study in this section, its assessment will be done in the following Chapter 6.

5.3 POLICY INTERVENTIONS IN BASIC EDUCATION

5.3.1 Case study 3: The impact of the introduction of Grade R on learning outcomes (Van der Berg et al., 2013)

The case presentation as detailed below is drawn from the above case document and all reference are drawn from the same case document.

5.3.1.1 Case background

The study set out to establish the impact of exposure to Grade R programmes on future learning outcomes of learners. From 2001, government implemented a compulsory 'reception year' for five-year-olds into the schooling system with a view to induct and prepare learners for entry to Grade 1 and the formal primary school system. Grade R was envisaged as an early childhood educational intervention to enhance children's cognitive skills towards assimilation into Grade 1 of primary schooling. The Grade R programme was formally introduced in all government primary schools as well as early childhood development centres. Previous studies were criticised for not providing conclusive evidence on the impact of Grade R as they studied general preschool attendance (Van der Berg et al., 2013: 39).

This study, using an alternative dataset, was commissioned to address this gap with a view to provide stronger evidence and causal links between attendance of Grade R and learning outcomes. The study was commissioned by the Department of Performance Monitoring and Evaluation (DPME) in the Presidency in partnership with the National Department of Basic Education (DBE) and undertaken by Stellenbosch University.

5.3.1.2 Scope of the evaluation

The scope encompassed 18 102 schools, of which 76.4 percent were primary schools, 20.2 percent combined schools and the remaining 3.4 percent intermediary schools across all nine provinces of the Western Cape (1 169), Northern Cape (407), Free State (992), Eastern Cape (4 772), KwaZulu-Natal (4 222), Mpumalanga (1 323), Limpopo (2 605), Gauteng (1 551) and North West (1 061) (Van der Berg et al., 2013: 14-15).

Schools studied were from quintile 1 to quintile 5, where the quintile system is a grouping of schools from poorest to least poor in terms of allocation of government subsidy to the school. Quintile 1 and 2 are the poorest schools that pay no fees, while quintile 5 are the wealthiest, pay fees and receive

the minimum government subsidy. Schools were also studied in terms of those that offer Grade R to Grade 6. Early childhood development centres that also offer Grade R were out of scope (Van der Berg et al., 2013: 15).

5.3.1.3 Objectives of the evaluation

According to documentary analysis, the question that the impact evaluation sought to answer was "whether it is possible to discern a causal effect of the introduction of Grade R on subsequent learning outcomes in school, and in particular the impact on children from disadvantaged home backgrounds" (Van der Berg et al., 2013: 4).

5.2.1.4 Hypothesis and theory of change

The theory of change should encompass context-mechanism-outcome suppositions. In this case, the theory of change for the intervention, or for the grade R evaluation was not found to be explicit. Nor was a hypothesis made about what might work for whom and in what circumstances for this particular intervention.

However, it was stated, "one of the aims of this reception year was to reduce the backlog faced by many learners in poorer schools due to a deficient home environment. The view was that providing support to such children at early ages should assist to reduce the backlogs they face when entering Grade 1" (Van der Berg et al., 2013: 4).

5.3.1.5 Impact evaluation methodology

The methodology employed was aimed at "measuring the impact of Grade R provision on learner performance in South Africa using a (proxy) measure of 'treatment', that is, the proportion of learners in a given grade in a given school that attended Grade R" (Van der Berg et al., 2013: 43). In this regard, the study applied a non-experimental quantitative method in the form of regression analysis modelling (Van der Berg et al., 2013: 14). The methodological model measured the correlation between Annual National Assessment (ANA) test performance in mathematics and home or first language against the key 'treatment' variable, i.e. the attendance of Grade R. Other independent explanatory variables included controls for the year of ANA testing (2011 or 2012), the grade of the student and various school characteristics. The size and significance of the estimated coefficient on the treatment variable represents the impact of having attended Grade R.

5.3.1.6 Data collection

Source data that informed the impact assessment came from three data sets:

- The Annual National Survey of Schools called SNAP, which had information on learner numbers in each grade across South Africa;
- The Annual National Assessments (ANA) of 2011 and 2012 that provided test performance in mathematics and home language for Grades 1 to 6; as well as

- Master data from the education information system (EMIS) which provided information on provincial and district location of the school, whether the schools were independent or public, the phase of the school, their quintile and the fees charged by the schools (Van der Berg et al., 2013: 41).

5.3.1.7 Data collected on context

A literature review providing background on both international and South African context focusing on early learning interventions within preschool education was conducted with a view of gaining insights on the effect of these interventions on children's cognitive, social and economic outcomes. Data was collected on schools in terms of their provincial and district location, whether the schools were independent or public, the phase of the school, their quintile and the fees they charged. Data about learners was also collected in terms of their numbers, grade levels and test performance (Van der Berg et al., 2013: 17-24).

5.3.1.8 Data collected on mechanisms

Mechanisms refer to the ways in which change occurs. The case study sought to find out whether the intervention works or not, in this instance the impact of Grade R on learning outcomes. These measured learning outcomes were in terms of cognitive abilities and skills, because of having attended Grade R. The gathered data did not show how that change takes place. The underlying reasons leading to achievement of learning outcomes were not explicit. Therefore, the mechanisms of the intervention in terms of how and why it worked for some schools and why it did not work for others were missing.

5.3.1.9 Data collected on outcome patterns

The study found that having attended Grade R, learners on average improved their mathematics score by 2.5 percent and language proficiency by 10.2 percent (Van der Berg et al., 2013: 16). Therefore, learners, from Grade 1 to Grade 6, who had attended Grade R, improved their mathematic test scores by 2.5 percent or by 12 days of learning in an academic year, as well as improved their language proficiency by 10.2 percent or by 50 days in an academic year. It was found that this effect was substantial for the higher quintiles schools and almost zero for low quintile schools and the effect was also stronger in Gauteng, Northern Cape and Western Cape provinces (Van der Berg et al., 2013: 23-24). The magnitude of the impact is significant when comparing high-performing provinces and weak-performing provinces and the school quintiles within those provinces. The impact of Grade R on mathematics outcome in a weak-performing province at lower quintile schools is very small at 1.8 percent compared to the same lower-quintile school in a high-performing province at 10.4 percent. Higher-quintile schools' mathematics performance in both weak-performing provinces and high-performing provinces is at 9.6 percent and 16 percent respectively. Therefore, learners in high-performing provinces, in both low-quintile and high-quintile schools, are better off in having successful learning outcomes than their counterparts in weak-performing provinces. Learners

in weak-performing provinces achieve better learning outcomes if attending Grade R in a higher-quintile school.

Having presented the case study in this section, its assessment will be done in the following Chapter 6.

5.4 POLICY INTERVENTIONS IN SOCIAL HOUSING

5.4.1 Case study 4: An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa (RSA, 2011b)

The case presentation as detailed below is drawn from the above case document and all reference are drawn from the same case document.

5.4.1.1 Case background

In South Africa, informal settlements became prominent in the 1970s and early 1980s when *Apartheid* influx control laws became lax and urbanisation levels rose as migrant labour came to major metropolitan cities in search of jobs and livelihood. *Apartheid* spatial planning did not cater for the housing needs of Africans and these informal settlements, erected by migrant labour as dwellings, mushroomed to large proportions in the intervening years.

In an attempt to upgrade and formalise housing, the government, amongst various initiatives, implemented the Upgrading of Informal Settlements Programme (UISP) project. According to case analysis (RSA, 2011b: 6):

> *…the main aim of this programme is to facilitate the structured incremental upgrading of informal settlements in cases where this is possible. Where this is not deemed possible, and as a last resort, the programme includes cases where communities must be relocated. Its main aims are to promote tenure security, health and welfare and community empowerment amongst those residing in informal settlements.*

The study was commissioned by the National Department of Human Settlements (NDHS) and conducted by the World Bank. The impact evaluation sought to find the causal links between the life circumstance of beneficiaries whose informal dwellings were either relocated to new houses or upgraded to serviced dwellings. According to the case study (RSA, 2011b:1), in Limpopo province an area known as Disteneng East had a total of 1 171 informal settlements and from these a treatment group of 444 households were relocated to Disteneng West to fully-serviced and formalized, subsidised Reconstruction and Development Programme (RDP) homes. The remaining 727 households in Disteneng East became the control group and remained in the same conditions as the treatment group left behind.

In another project in the Free State province, the original informal settlement was upgraded. In an area known as Grassland, outside Bloemfontein 1 014 households were selected. Of these, 659 households were provided with housing without sanitation, i.e. the treatment 2 group. Residents

from adjacent Bloemhof making part of the sample at 355 were provided with serviced stands that had running water, electricity and sanitation but without the housing, i.e. the treatment 1 group (RSA, 2011b: 1).

In Gauteng province in the Chris Hani Settlement in Daveyton, a sample of 1 303 households was selected. The informal settlement consisting of three extensions was undergoing an upgrading programme of houses, electricity and sanitation. Extension 3 consisting of 398 households was extensively upgraded to a house including water, sanitation and electricity. Extension 3 was compared to Extension 1 and 2, totalling 905 households, which had partial upgrading of houses, water and sanitation, excluding electricity (RSA, 2011b: 1).

5.4.1.2 Scope of the evaluation

The scope of the evaluation was limited to defined informal settlement households and newly-built RDP household sites in the Free State, Gauteng and Limpopo provinces (RSA, 2011b: 1).

5.4.1.3 Objectives of the evaluation

The analysis garnered in the case illustrated the stated objectives as “to rigorously measure the impact of the UISP on the welfare of local communities across a broad range of indicators, and (in future rounds of the study) to investigate whether specific interventions or combinations of interventions are more cost-effective than others in achieving positive outcomes” (RSA, 2011b: 7).

In addition, the human settlement plan, Breaking New Ground Strategy describes these lifestyle outcomes to be promoted as income levels, employment, investment, health, savings and child development (RSA, 2011b: 12). The case focused on the first objective, as the second objective will be embarked upon in later impact evaluations.

5.4.1.4 Hypothesis and theory of change

Case study analysis did not find that a hypothesis was made about what might work for whom and in what circumstances for this particular intervention. Case study analysis did not illustrate a clear and explicit theory of change showing causal linkages between informal settlement upgrading and the life circumstances of beneficiaries whose informal dwellings were either relocated to new houses or upgraded to serviced dwellings. Therefore, there was no specific programme theory that was elicited, formalised and tested.

5.4.1.5 Impact evaluation methodology

The study employed a natural experiment, quasi-experimental methodology to determine causal impact between relocation from an informal settlement, *in situ* upgrading of an informal settlement and partial upgrade of an informal settlement against a number of well-being indicators (RSA, 2011b: 27-29). Participants in the study naturally became allocated into either ‘treatment’ and ‘control’ groups due to chance events or the housing service delivery circumstances. This then informed the natural experiment of the study.

5.4.1.6 Data collection

Trained enumerators administered a household questionnaire consisting of 14 sections. The questionnaire measured various aspects. These were household demographics, education levels, economic activity of the household, healthcare, and savings and credit patterns of the household. I addition, small enterprise activity of the household, incidence and levels of crimes, housing tenure, the quality and accessibility of service delivery and infrastructure were measures as well as social and community participation, level of satisfaction with municipal services, the neighbourhood, and improvements in physical conditions of the household (RSA, 2011b: 36).

5.4.1.7 Data collected on context

Data collected focused largely on household demographics, educational levels in terms of literacy rates, school enrolment and attendance and pass rates. Other household characteristics focused on incidence and severity of disease and injury of each member, the economic activity of the households in terms of small business activities and jobs. Access to credit and savings patterns were also interrogated. The external setting surrounding the home was also interrogated; these included the impact of crime and violence, infrastructure and service delivery and the social environment (RSA, 2011b: 36).

5.4.1.8 Data collected on mechanisms

Mechanisms refer to the ways in which change occurs. The case sought to determine causal impact between relocation from an informal settlement, *in situ* upgrading of an informal settlement and partial upgrade of an informal settlement against a number of well-being indicators (RSA, 2011b: 27). The gathered data did not illustrate any evident mechanisms of the intervention in terms of how and why it worked.

5.4.1.9 Data collected on outcome patterns

The outcomes envisaged in the objectives were improvements in income levels, employment, investment, health, savings and child development (RSA, 2011b:12). It was found that relocation from an informal settlement to a house had an insignificant impact on income. Findings indicated that the average monthly household income of R1 501 in the control group did not differ significantly from R1 632 in the treatment group. Partial upgrade increased income from R988.13 to R1195.88. In addition, the per capita income was R536 in the treatment group, significantly less than R999 found in the control group (RSA, 2011b: 60).

In terms of employment levels, it was found that after relocation, the broad unemployment rate rose from 42 percent in the control group to 56 percent in the treatment group, and when looking specifically at the household head, the unemployment rate rose from 31 percent to 48 percent (RSA, 2011b: 65). When informal settlements were partially upgraded, 52 percent of households had at least one-person unemployed (RSA, 2011b: 138).

It was found that relocation to a house had a positive impact on property investment. Households were 15 percent more likely to invest in their properties spending on average R999 on home improvements as well as 12 percent more likely of planning to use their savings for home improvements than the control group. In the Grassland project, households spent on average R1 463 on home improvements, given that they had a house without sanitation. Households were also found to increase rental income, from backyard tenants, to R468 per month compared to the control group at R303 per month. Rental income is a major (RSA, 2011b: 52-75). Relocation to a house was found to have a positive impact on children's health as morbidity dropped from 40 percent to 25 percent. This was attributed to a more hygienic home environment.

Having presented the case study in this section, its assessment will be done in the following Chapter 6.

5.5 SUMMARY AND CONCLUSION TO CHAPTER 5

The chapter presented the micro-analyses of four case studies. The first two cases were policy interventions on social protection namely, *The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households* and the *Youth wage subsidy experiment for South Africa.*

These were followed by a case study on policy intervention in basic education, *The impact of the introduction of Grade R on learning outcomes* and a case study on a social housing policy intervention, i.e. *An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa*

The aim of conducting analyses of these case studies was to establish the methodologies and approaches used in past programme impact evaluations within the South African public sector as well as establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.

The micro-analyses applied the Realist Evaluation Impact Evaluation Assessment Framework as a conceptual framework to evaluate each case study in search of the purpose of the evaluation, the scope and objectives and the impact evaluation methodology applied, an explicit theory of change as specified in the hypothesis to establish how the programme works, in what context and under what conditions, confirming if data was collected on appropriate mechanisms, contexts and outcomes. After this, each proposed CMO configuration was analysed to see if it is supported by the observed programme outcome patterns in terms of the findings. Finally, the analyses checked that the initial hypotheses were in fact true, and if not, whether they were further refined and reframed towards alignment with observed outcome patterns. The Realist Evaluation Impact Evaluation Assessment Framework was applied in these four case studies to test the robustness of the methods used in past impact evaluation in offering policy-makers the most useful findings.

The next chapter discusses the findings from the key informant interviews with policy decision-makers and commissioners and implementers of public sector evaluations.

CHAPTER 6:
INTERVIEW FINDINGS FROM POLICY DECISION-MAKERS AND COMMISSIONERS OF IMPACT EVALUATIONS

6.1 INTRODUCTION

This section presents the data from the key informant interviews. Within the overall aims of the study, the objectives of the key informant interviews were two-fold. First, to establish from policy decision-makers, commissioners and implementers of evaluations in the South African public sector the utility value of evaluation results in offering new insights of what works, for whom, under what conditions and in what respects which can possibly result in evaluation findings that are meaningful, have utility value and aid in policy-making. Secondly, to establish from the same respondents the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector

6.2 THE RESEARCH INTERVIEW FINDINGS

The overall focus and purpose of the questions posed was to find out from the respondents what makes evaluation findings, meaningful, valid and useful. This was in line within the overall research context of establishing the potential value of Realist Evaluation results in offering insights of what works, for whom, under what conditions and in what respects which can possibly result in evaluation findings that are meaningful, have utility value and aid in policy-making.

To further validate and triangulate the data from the key informant interviews, parliamentary questions and answers raised by policy-makers subsequent to these impact evaluations are also provided as further input substantiating the key informant interviews.

6.2.1 What is expected from policy evaluations

Respondents were provided with a list describing aspects of policy evaluation and were asked which aspects from this list do they most expect or need. Table 6.1 illustrates the list provided to respondents including the respondents ranking of these aspects. The full interview instrument may be found as Appendix E.

Percentages are used as a means of facilitating the ranking of preferences rather than implying significance of the population, as the population was too small. The responses are graphically demonstrated in Figure 6.1 and discussed below. All respondents at 100 percent, expected policy evaluations to be coherent in specifying the outcomes and results of an evaluation. Explanations in this regards were that “these are most important to understand the possible route to change” (Commissioner Education Evaluations, 2016) as well as “these provide a link with policy objectives (Commissioner #2 Social Development Evaluations, 2016).

Table 6.1: List describing aspects of policy evaluations

No.	Ranking	Aspects
1.1	1	What were the outcomes / results of the policy evaluation
1.2	2	The problems encountered in implementing the policy
1.3	3	How the policy was implemented
1.4	6	Any unexpected programme outcomes
1.5	2	Who benefited primarily from the policy
1.6	2	If the intervention was successful, when and where can it be replicated
1.7	6	The impact of demographic effects like culture on the expected programme outcomes
1.8	7	Who did not benefit from the policy
1.9	5	Why some do and others do not benefit from the policy
1.10	4	The theory assumptions on how change is supposed to happen
1.11	6	The budget

Source: Author.

The second-most highly-rated aspects of policy evaluation at 86 percent each focused on the specification of policy implementation problems, who primarily benefited from the intervention as well as when and where the intervention can be implemented if successful. Insights on implementation problems were seen as critical in order to improve the programme design. Some indicated that insights into these implementation problems and issues are important because “These gives us an idea of why intended outcomes are not reached, for example the requirement of a birth certificate has been a hindrance to accessing the CSG” (Commissioner #1 Social Development Evaluations, 2016).

Two respondents indicated that the problems had to be known “in terms of where the breakdowns occur in the pathways to change” (Sector Expert Human Settlements Evaluations, 2016) since these “insights feed into rectifying programme implementation and administrative effectiveness in line with intended outcomes” (Commissioner Education Evaluations, 2016). Who exactly end up benefitting from the policy intervention was equally important. It was deemed that “after quantifying who gets and who does not, factors leading to uptake and non-uptake are critical to understanding how the programme works” (Commissioner Education Evaluations, 2016). This was also important in terms of equity as “who exactly benefited and how equity is dispersed towards impact is critical in order to know whether the targeted beneficiaries were indeed the beneficiaries (Sector Expert Human Settlements Evaluations, 2016). Therefore, policy decision-makers want to know “whom did it benefit and how did it work. To know whether policy should be targeted or done at full scale.” (Sector Expert Education Evaluations, 2016)

The possibilities of implementing the programme in other settings was seen as important because “replication for policy-makers is very important. Can this be replicated elsewhere? Could it work elsewhere? You want to know the average effect in order to improve policy design” (Sector Expert Education Evaluations, 2016).

Third most rated aspect was “how the policy was implemented” This was supported by the views of approximately two thirds or 71 percent of the respondents who also expected an evaluation to specify how it was implemented.

The fourth aspect was an expectation of the evidence of a theory of change in policy evaluation from at least 57 percent of the respondents. This was critical because “the ToC tells the impact of the programme, how it works, its pathways and indicates the critical assumptions.”

The fifth aspect was an expectation to know why some do and some do not benefit from the policy mentioned by 43 percent of the sample. The sixth aspect focused on information on unexpected outcomes, the impact of demographic effects like culture on the expected programme outcomes as well as the budget of the evaluation. These were mentioned by 29 percent of the respondents respectively. The seventh and last aspect was an expectation for the policy evaluation to indicate those who did not benefit from the policy and this was rated by 14 percent of the respondents.

In conclusion, the commissioners and implementers of evaluations expect to derive value from policy evaluations that clearly specify the outcomes and results of an evaluation; articulates the implementation problems in order to improve policy design; are coherent about the equitable nature of the policy in terms ascertaining if the intended beneficiary did benefit as well the implications for programme applications in other settings. In addition to these factors, insights on programme pathways should be specified in policy evaluations as well as the observed unintended outcomes of a policy evaluation. Budgetary issues regarding the cost of the evaluation, though to a much smaller extent is also somewhat expected.

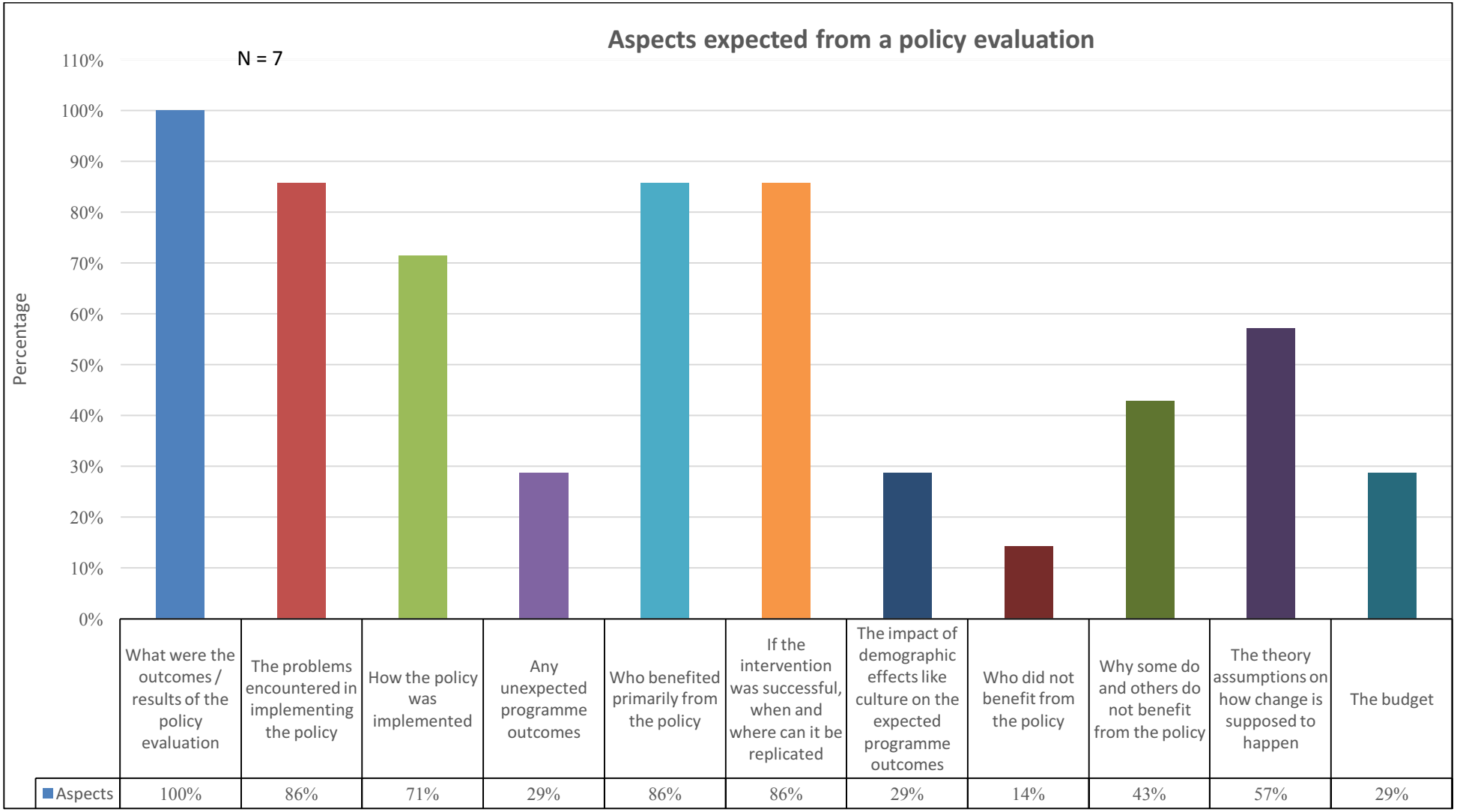


Figure 6.1: As a policy-maker, which of the following aspects do you 'most expect/need' from a policy evaluation? Please tick all that is relevant to you

6.2.1 The high priority aspects of policy evaluations

After having highlighted what is expected or needed in policy evaluations, the respondents then ranked these in order of priority, highlighting the top five and most critical aspects that will result in the most meaningful evaluations. In the initial previous question, the general aspects of what is expected were indicated. In this question those same aspects were ranked into the five most critical aspects whose presence or absence strengthen or weaken an evaluation. These results are highlighted in Figure 6.2 below.

The most critical and important aspect of an evaluation is “What were the outcomes / results of the policy evaluation”. This was highly prioritised by approximately 86 percent of all respondents. The respondents were asked to justify their rankings in terms why they think policy makers would regard this aspect to be most important. It was highlighted that “the results of outcomes are important and very primary” (Sector Expert Education Evaluations,2016). Also “outcomes provide a link with policy objectives” (Commissioner #2 Social Development Evaluations, 2016), as well the fact that “outcomes are critical” (Policy Expert Social Development Evaluations,2016). In addition, “outcomes yield broad success and failures of the programmes” (Sector Expert Human Settlements Evaluations, 2016).

The "theory assumptions on how change is supposed to happen” was prioritised as the second most critical aspect of an evaluation by 71.4 percent of the respondents. Whilst information regarding the programmes’ theory of change assumptions was a fourth expected aspect in an evaluation in the previous question, when it comes to prioritising the five most critical aspects that add credibility to an evaluation, clarity on theory of change assumptions are highly prioritised by decision-makers. This may primarily be the result of the value placed on articulating and understanding the programme’s pathways to change. This has been supported by the literature which highlighted that the use of programme theory is now conventional and is a core requirement as evaluation commissioners require project proposals to initially specify the theory of change as a guide for assisting in programme design and evaluation (Rogers, 2007:63-64). The respondents were asked to justify their rankings in terms of why they think policy makers would regard this aspect to be most important. It was highlighted by some that “ToC tells the impact of the programme and how it works and provides insights on how and why this programme should work to its intended outcome” (Sector Expert Human Settlements Evaluations, 2016). Others indicated that “these are most important in order to understand the possible route to change” (Commissioner Education Evaluations, 2016).

In the third place, “The problems encountered in implementing the policy” was highlighted by 57.1 percent of the respondents. This was deemed important because this offered “insights into rectifying of programme implementation and administrative effectiveness versus the intended outcomes” (Commissioner Education Evaluation, 2016). Whilst another respondent was of the opinion that this “gives an idea of why intended outcomes are not reached as implementation issues are important.

For example, Home Affairs requirement of birth certificate poses a hindrance to the CSG" (Commissioner #1 Social Development Evaluations, 2016)

"Who benefited primarily from the policy" occupied the fourth place ranked by 43 percent of the respondents. This was because issues on "whether the target group is being reached such as through the Pro-Poor Policy" (Commissioner #2 Social Development Evaluations, 2016) as well as "whether the targeted beneficiaries were indeed the beneficiaries which addressed issues of equity" was important (Sector Expert Human Settlements Evaluations, 2016).

Lastly, on the fifth position, "How the policy was implemented" was prioritised as critical by 29 percent of the respondents. This was deemed important because it was important to gain "insights into how to scale up especially in terms of intended and unintended aspects of implementation (Commissioner Education Evaluation, 2016), in addition "policy implementation and programme design is important to facilitate review. Its about learning" (Commissioner #2 Social Development Evaluations, 2016).

The researcher further asked the respondents whether they think all policy-makers would agree with the rankings they offered. The majority of respondents agreed that policy-makers would most probably agree with their respective rankings except for one respondent who indicated that "if someone is new to the party they won't agree. For example, the Nutrition Programme ToC was not understood by people new to the evaluation. They won't rank the same but would agree with issues as influenced by their institutional memory of the programme. Those who started with the programme will have better understanding of the critical aspects of the programme" (Commissioner Education Evaluations, 2016).

Whilst information regarding programme replication was expected in an evaluation, it is not highly prioritised as evidenced by its absence in the top five critical aspects of credible and useful evaluations. This may be due to an appreciation by commissioners of evaluations that replication of social programmes in social settings is not always feasible. Social programmes are not laboratory experiments that can simplistically be replicated from one setting to another. It is possible that if the researcher framed this aspect using terms such as 'applied', or 'adapted to', the perception of the respondents could have been somewhat different.

One may deduce from this ranking that valid and useful policy evaluations should at minimum, clearly indicate the outcomes and results of the evaluation and provide a coherent programme theory of change that indicate how the programme results in the desired change. Following this, such evaluations should highlight any implementation challenges in order to improve programme design and enhance programme efficacy. Equally critical, the evaluation should provide assurance that the observed outcomes are equitable in reaching the targeted beneficiaries as well mapping out the policy implementation process.

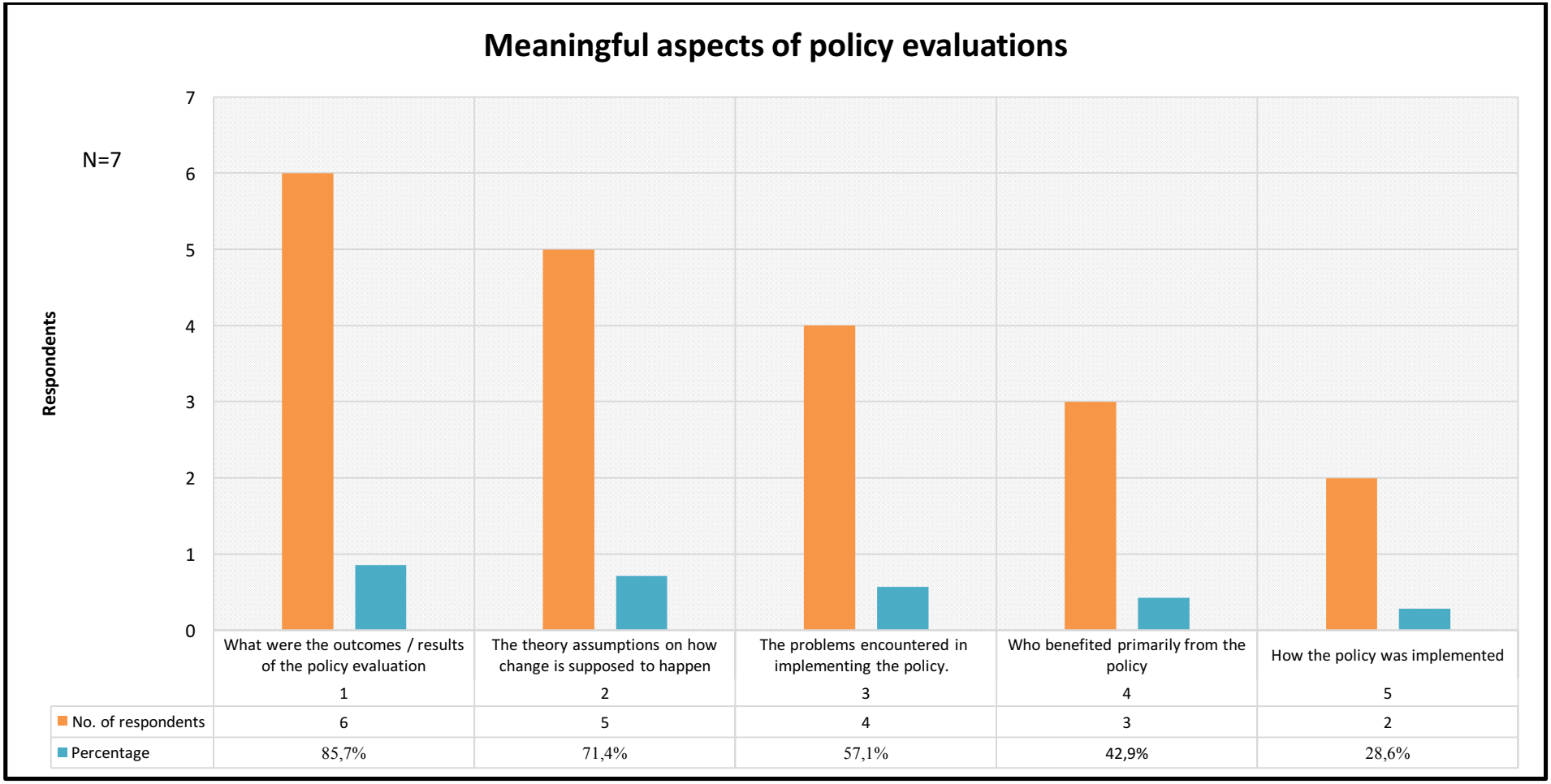


Figure 6.2: What policy-makers deem to be the 5 most meaningful, valid and useful important aspects in policy evaluation

6.2.2 The most important gaps and limitations with existing policy impact evaluations

With the overarching aim of understanding from policy decision-makers, commissioners and implementers of evaluations, the factors that make evaluation findings meaningful, valid and useful, respondents were asked to reflect on their experience and identify the most important gaps/limitations with existing policy impact evaluations. Key themes and patterns were identified during data analysis; these are reflected in Table 6.2 and were highlighted as:

- Evaluation methodologies and designs;
- Lack of programme theory of change;
- Inadequate evidence-based policy-making; and
- Budgetary constraints.

The first prominent recurring theme focused on the evaluation methodologies and designs. Two experts in the education sector, and one in social development agreed that impact evaluation methodologies and the way evaluations are designed posed critical limitations. The methodological approaches were found wanting, by all three when used for the assessment of attribution and causality, and ultimately finding out ‘what works’.

A second theme that came out from the data analysis was the absence of the theory of change that establish how the programme works, in what context and under what conditions as well as the clear demonstrations of programme impact. This was most highlighted by one expert in social development sector and one other expert in the human settlement sector.

A third theme was limitation in the use of evaluation evidence in policymaking. A key factor highlighted in this regard was the limited utilisation of evaluation evidence whether to inform or as a basis for policy-making. It was highlighted that the policy cycle often progresses without diffusing the available evidence into policy-making. These views were highlighted by two experts in the education sector, one expert in social development sector and one expert in the economic cluster.

A fourth and final theme regarding the most important gaps/limitations with existing policy impact evaluations was the issue of public sector budgetary constraints which influenced whether, and what type of evaluations are actually completed. These views were highlighted by three experts, two of whom were from the social development sector and one from the education sector.

Consequently, expert commissioners and implementers of evaluations in the public sector have indicated that there are important gaps and limitations with existing policy impact evaluations. Evaluation methods and designs are not always appropriate to inform the needs of policy-makers. There is limited insights on programme pathways to change as a base of establishing how the programme works, in what context and under what conditions. There is also perceived limited utilisation of evaluation evidence in policy-making, as evaluation evidence is not effectively infused in the policy-making cycle. Finally, resource constraints in the public purse also have a bearing on whether, and what type of evaluations are actually accomplished.

Table 6.2: Limitations with existing policy impact evaluations

Limitations with existing policy impact evaluations	Illustration from the data
Adopted evaluation methodologies and designs	"A lot of evaluations don't say anything about impact. They are not designed as impact evaluations. They simply do not have the data or the design to be able to conclude what was the cause or impact of programme on intended beneficiaries. They are not designed that way" (Sector Expert Education Evaluations, 2016).
	"The same people do evaluations in the public sector. As a result we get same evaluation approaches and same methodologies" (Policy Expert Social Development Evaluations, 2016).
	"We don't know what affects uptake of treatment. We do not know very much. We can harvest underlying qualitative aspects. This is becoming more prevalent. We try to explore the why of uptake. We don't know what the uptake issues are" (Commissioner Education Evaluations, 2016).
	"We can speculate, look at implementation, but we don't have studies on impact, it is because programmes operate at scale and there is no obvious counterfactual to look at" (Sector Expert Education Evaluations, 2016).
Lack of programme theory of change	"Most programmes do not have a Theory of Change as the design lacks it" (Commissioner #2 Social Development Evaluations, 2016).
	" The Theory of Change is missing because what is it that we want to know and why is missing. Whether this is an appropriate methodology to answer that or whether another research tool could better serve that is not clear(Sector Expert Human Settlements Evaluations, 2016).
	"Policy-makers set out interventions but have not been able to assess failure or success because RCT's are broad. We need ToC to make sense of programmes. The "why" is missing as M&E units do not have skills to do this. Successful evaluations are a combination of methodology and ToC in a specific sector" (Sector Expert Human Settlements Evaluations, 2016).
Inadequate evidence-based policy-making	"We don't have impact evaluation evidence. This goes for effect of textbooks on learning outcomes, scholar transport programmes, teacher training. Impact on outcomes is missing on these" (Sector Expert Education Evaluations, 2016).
	"Evidence is not used for policy-making because of how Government works and the way policies are structured. Because of this, you have to implement immediately without the benefit of an evaluation. Government does not rely on evidence" (Commissioner #1 Social Development Evaluations, 2016)
	"We often don't know if the programme works on average e.g. school feeding programme, its impact on learner attendance, retention at school, cognitive outcomes and nutritional outcomes are unknown" (Sector Expert Education Evaluations, 2016)
	"There needs to be a paradigm shift towards various policies on a small scale before roll out as part of evidence-based policy-making. Challenge is that government does not want evidence because political patronage is taken away in evidence-based policy-making. Policy goes ahead regardless of whether there is evidence or not" (Policy Expert Economic Cluster Evaluations, 2016).
	"We need to know about policy-makers themselves and their own incentives to use information. Some policy-makers by virtue of their backgrounds are more receptive to the use of information, others not as receptive" (Commissioner Education Evaluations, 2016).
Budgetary constraints	"It is about who determines the agenda in terms of money, budget and political clout" (Policy Expert Social Development Evaluations, 2016)
	"Budget is critical. Government does not always have money to allocate to evaluations. The current Early Childhood Development budget determines what evaluation approach and methodology will be adopted" (Commissioner #1 Social Development Evaluations, 2016).
	"Relative cost to evaluate is also important. Cost-benefit is important (Commissioner Education Evaluations, 2016).

6.2.3 Benefits of past evaluations to policy-makers

It was further established whether the findings from the evaluations that the respondents were intimately involved with, did contribute meaningful findings to policy-makers as well as whether there were some important gaps and limitations. In this regard, the respondents were asked to reflect on the most valuable contributions of the evaluation findings to policy decision-makers, as well as their opinions on the most important gaps/limitations of those evaluations to policy decision-makers.

The results are presented per case study, starting with the positive contributions and benefits derived from the evaluation for each case. This is then followed by the shortcomings and limitations of each case.

6.2.3.1 CSG contributions

- The evaluation served to inform the NEPF (Commissioner #2 Social Development Evaluations, 2016).
- The evaluation had a positive impact on children's lives (Commissioner #1 Social Development Evaluations, 2016).
- The evaluation identified which groups were not benefiting, which served to illustrate the importance of early exposure to avoid later risky behaviour (Commissioner #1 Social Development Evaluations, 2016).
- The evaluation served to strengthen the ECD policy and informed the development of a more comprehensive ECD policy (Policy Expert Social Development Evaluations, 2016).
- The evaluation informed policy option regarding universalising the CSG (Commissioner #2 Social Development Evaluations, 2016).

6.2.3.2 CSG shortcomings

- The evaluation had methodological inadequacies such as the absence of a proper control group as randomised controlled trials were not appropriate in this context (Commissioner #1 and Commissioner #2 Social Development Evaluations, 2016).
- The evaluation did not address all the needs of policy-makers. Appendix F lists questions that were not clear following the evaluation
- The evaluation was very costly (Commissioner #2 Social Development Evaluations, 2016).
- The expertise to undertake the evaluation was outside government (Commissioner #1 Social Development Evaluations, 2016).

6.2.3.3 Youth wage subsidy contributions

- It was found that tax credit policies have an impact in smaller companie(Policy Expert Economic Cluster Evaluations, 2016).
- The evaluation informed policy choices as in 2014, the wage subsidy was implemented through the tax system as a tax incentive policy (Policy Expert Economic Cluster Evaluations, 2016). Government announced that the proposed youth employment incentive would be run through the South African Revenue Service (SARS) as a tax incentive to firms that hire young, inexperienced workers. The estimated cost to government was approximately R5 billion over three years through foregone tax revenue. The actual amount was dependent on firm uptake and job creation as result of the subsidy (Parliamentary Monitoring Group, 2012b).

6.2.3.4 Youth wage shortcomings

- The evaluation did not address all the needs of policy-makers. Appendix G lists questions from policy-makers that were not clear following the evaluation.
- The evaluation did not address all the needs and concerns of organised labour as organised labour had stronger views against the wage subsidy. In addition, issues raised by organised labour were not fully engaged (Policy Expert Economic Cluster Evaluations, 2016).

6.2.3.5 Grade R contributions

- The evaluation demonstrated a plausible estimate of causal impact on those that attended Grade R and its effect on learning outcomes in later life (Sector Expert Education Evaluations, 2016).
- The evaluation informed policy options regarding quality of teaching of Grade R, the quality of learning outcomes and the allocation of resources such as the remuneration of Grade R teachers (Commissioner Education Evaluations, 2016).
- The evaluation highlighted some inconsistencies in the provision of Grade R and provided limited understanding of what was working and not working (Sector Expert Education Evaluations, 2016).

6.2.3.6 Grade R shortcomings

- The evaluations had limitations as it was a desktop review and issues of external validity, generalisability and legitimacy were raised as “no fieldwork was done and Grade R classrooms were not visited” (Sector Expert Education Evaluations, 2016).
- The evaluation did not address all the needs of policy-makers. Appendix I lists questions from policy-makers that were not clear following the evaluation

- The design and methodology of the evaluation posed limitations as critical factors in early childhood education were not part of the evaluation (Policy Expert Social Development Evaluations, 2016).

6.2.3.7 Upgrading of informal settlements (UISP) contributions:

- The evaluation highlighted the importance of the right combination between housing financial structuring and human settlement planning instruments which when implemented effectively can affect pathways to change (Sector Expert Human Settlements Evaluations, 2016).

6.2.3.8 Upgrading of informal settlements (UISP) shortcomings:

- The evaluation did not address all the needs of policy-makers. As demonstrated in Appendix J, policy-makers had a list of questions for which the answers were not clear following the evaluation.

6.2.4 Potential value of Realist Evaluation approach in policy impact evaluations

This question was preceded by a brief explanation of Realist Evaluation approach by the interviewer, highlighting the potential value of the approach on impact evaluations. The Realist Evaluation theoretical concepts were briefly explained including the nature of evaluations findings from the approach. The respondents were then asked if they thought adopting such an approach would offer useful, meaningful or valid results to policy decision-makers, perhaps beyond what they derived from the existing evaluations they were part of. All respondents agreed that such an approach can potentially add value to evaluation findings.

Whilst all respondents saw potential value in the application of the Realist Evaluation approach, two recurring themes were highlighted, namely the systematic and methodical nature of the approach as well as its interrogation of programme context. These aspects were seen as the two primary value adding factors of the approach.

Realist Evaluation was regarded as valuable given its theory-based foundations that inform systematic analysis. Experts in the social development sector were of the opinion that “it depends on what you want to achieve. However, one can effectively unpack issues through this approach.” (Commissioner #1 Social Development Evaluations, 2016). Views from the human settlement sector were that the REM’s theory-based underpinnings can be effective in offering insights on specific sub-groups, as “it can help us look at sum of pathways and can help us better disaggregate housing needs as it is good to have tailored responses to tailored sets of situations. The value of an evaluation is an evidence base with good theory. Realist synthesis is excellent in this and therefore its value is obvious” (Sector Expert Human Settlements Evaluations, 2016). Others valued the potential insights of learning about the specific context and conditions that can influence a programme to work most effectively. In this regard, one expert stated “Realist Evaluation has the potential of providing more

depth and not leave things at superficial level. The why and how it worked is important. I would love that approach. I would want to know if it worked in this context, but why?" (Policy Expert Social Development Evaluations, 2016)

Equally, considering this, another expert also stated that "people are not homogeneous groups. There are differences within them. Hence, you might want to find out under which circumstances this impact will occur. This approach can offer more value to policy-makers." (Commissioner #2 Social Development Evaluations, 2016).

Similarly, another expert in the education sector indicated that adopting REM could have enhanced previous evaluation results, perhaps beyond what was derived. "Realist Evaluation would have added value to this study. It would definitely have helped. These are important questions. We are interested in knowing when is measurable impact valid. Does a type of evaluation influence it? In this regard theory-driven approaches can be helpful" (Sector Expert Education Evaluations, 2016).

On the other hand, other respondents identified potential drawbacks of adopting a Realist Evaluation approach, including:

- The detailed and cumbersome process of Realist Evaluation;
- The complex nature of theory-based evaluations;
- The supposed lack of consensus about the pathways to change and programme mechanism of public sector programmes which are taken as a given in Realist Evaluation;
- Lack of common understanding on programme input, processes and outcomes;
- Vast human and financial resources required to effectively carry out a Realist Evaluation;
- Impact evaluations currently not feasible in the public sector due to nascent nature of monitoring systems.

As a theory-driven approach, Realist Evaluation is meticulous, systematic and methodical in its research approach. The systematic and methodical nature of the approach could result "in a lengthy process as the usefulness of the approach depends on sample and depth of answers given since it requires open ended questions" (Commissioner #1 Social Development Evaluations, 2016). Concerning this, it was granted that whilst there is value in adopting the approach "theory-based evaluation needs multiple pathways and evidence. You could lose people in the detail. Realist Evaluation is demanding and interrogates in detail" (Sector Expert Human Settlements Evaluations, 2016). Furthermore, there is as yet no sectorial unanimous consensus on programme's theory of change within specific policy sectors of government. Adopting a Realist Evaluation would assume that this is in place which is not the case at present. Therefore, it was highlighted that "Realist Evaluation makes assumptions about theories of change. It assumes that we agree on pathways; however, there is no agreement about intervention logic. There is equally no agreement on mechanism" (Sector Expert Human Settlements Evaluations, 2016).For these reasons and due to lack of common understanding of programme pathways to change it was highlighted that "because

government is implementing and administering we have effectively decided to do implementation evaluations and then follow those up with impact evaluations in a few years by which time there is common understanding of programme input, processes and outcomes" (Commissioner Education Evaluations, 2016).

In addition to these factors, vast human and financial resources were required for en effective Realist Evaluation which may not be readily available. In this regard one expert from the education sector highlighted that the approach "could have added value to the Grade R evaluation but it could have been expensive because of detailed data collection methods" (Commissioner Education Evaluations, 2016). In addition to these factors it was indicated that adequate time is also required in order for change mechanism to be observable because "you can't implement and then quickly ask those complex and deeper questions because change and impact might not be there in the short term. Test of time is needed for rigorous impact evaluations" (Commissioner #2 Social Development Evaluations, 2016). Therefore, Realist impact evaluations may currently not be feasible in the public sector due to the nascent nature of monitoring systems.

Therefore, the findings indicate that whilst there is potential value in adopting a Realist Evaluation by virtue of its systematic and methodical nature, theory driven underpinnings and interrogation of programme context, this has to be balanced with other factors such as an understanding of the detailed methodical process by evaluators and commissioners alike, consensus by all stakeholders regarding what the programme pathways to change are within the specific policy sectors, enough lead time to effectively execute the evaluation as well as adequate financial resources to fund the evaluation.

6.2.5 The appropriateness of evaluation approaches in general in the public sector to inform the needs of policy decision-makers

The views of policy decision-makers, commissioners and implementers of evaluations were investigated in order to establish the usefulness of the evaluation results in offering new insights. The interviewed key informants indicated that the evaluations that are sometime applied in the public sector are not always appropriate to inform the needs of policy-makers. Key issues raised centred on:

- The types of evaluations that are currently done;
- Lack of attribution and causality; and
- Evaluation capacity issues.

Currently, the most prevalent type of evaluations currently carried out in the public sector are implementation and diagnostic evaluations. This is the view of two education experts in the indicating that, "attribution is a weakness. There is not enough data on programme outcomes. Currently it is mostly implementation and diagnostic evaluations that are done." (Sector Expert Education Evaluations, 2016), as well as that "implementation evaluations are currently being done and are a

key focus since there is not enough data to do impact evaluations. Implementation fidelity is not sufficient and consistent to be able to draw a counterfactual" (Commissioner Education Evaluations, 2016). Whilst this is the case, experts in the social development sector felt that progression had to be made towards the demonstration of attribution and causality and held that "the diagnostic approach that is being done is important but you need to start to do impact evaluations. Impact Evaluation must be done at a point to provide a clear roadmap" (Policy Expert Social Development Evaluations, 2016).

The absence of attribution and causality in programme evaluation was also echoed by an expert in the economic cluster, claiming that, "rigorous impact evaluations are required and are important for outcomes. The current approaches are not always appropriate. Within the economic cluster in government, generally evaluation is missing and done poorly." (Policy Expert Economic Cluster Evaluations, 2016). This lack of studies on impact was further acknowledged by experts in the education sector indicating that, "we can speculate and look at implementation, but we don't have studies on impact. It is because programmes operate at scale and there is no obvious counterfactual to look at (Sector Expert Education Evaluations, 2016).

As a result of these factors and lack of attribution and causality in programme evaluations, it was posited that policy-makers are not well positioned to effectively assess failure or success. In this regard one expert maintained that, "successful evaluations are a combination of methodology and theory of change in a specific sector. Policy-makers set out interventions but have not been able to assess failure or success. We need theory of change to make sense of programmes. The "why" is missing and M&E units do not have skills to do this (Sector Expert Human Settlements Evaluations, 2016).

A key finding from this study was that the expertise for the design of impact evaluations specifically for complex interventions is mostly outside the public sector. This raises the issue of evaluation capacity within the public sector. Impact evaluations that have been completed have largely been led by multinational expert teams who had the skills and know-how to design highly complex evaluations. Some views in this regard were that "evaluation designs are a problem. We need expertise since evaluations are still new in government. Currently evaluation expertise is not in government" (Commissioner #1 Social Development Evaluations, 2016). Other respondent echoed the same views highlighting for example that "the CSG evaluation was highly complex, it was the first rigorous impact evaluation in South Africa and international experts were borrowed from various countries. This posed challenges in terms of the management of the Terms of Reference as well high level coordination" (Commissioner #2 Social Development Evaluations, 2016).

Consequently, from the perspective of policy decision-makers who are responsible for commissioning evaluations and providing policy expertise, the appropriateness of evaluation approaches in general in the public sector pose some limitations. This is primarily influenced by specific types of evaluations that are currently dominant in the public sector. This has consequences

as programme attribution and causality, largely achievable through impact evaluations remain sparse. Because impact evaluations are not prevalent, the expertise to actually conduct them remains weak within the public sector. This results in a vicious circle as each of these factors reinforces each other.

6.3 SUMMARY AND CONCLUSIONS TO CHAPTER 6

The objectives of this section were to establish from policy decision-makers, commissioners and implementers of evaluations in the South African public sector what the factors are that add value in policy evaluations, which would then lead to better and new insights of what works, for whom, under what conditions and in what respects. In essence, what causes valuation findings to be meaningful, have utility value and aid in policy-making.

It was concluded that the commissioners and implementers of evaluations expect to derive value from policy evaluations that clearly specify the outcomes and results of an evaluation; articulate the implementation problems in order to improve policy design; are coherent about the equitable nature of the policy in terms ascertaining if the intended beneficiary did benefit as well the implications for programme applications in other settings. In addition to these factors, insights on programme pathways should be specified in policy evaluations as well as the observed unintended outcomes of a policy evaluation. Budgetary issues regarding the cost of the evaluation, though to a much smaller extent is also somewhat expected.

These factors were also ranked in terms of their priority and importance and it was concluded that valid and useful policy evaluations should at minimum, clearly indicate the outcomes and results of the evaluation and provide a coherent programme theory of change that indicate how the programme results in the desired change. Following this, such evaluations should highlight any implementation challenges in order to improve programme design and enhance programme efficacy. Equally critical, the evaluation should provide assurance that the observed outcomes are equitable in reaching the targeted beneficiaries. Policy decision-makers, commissioners and implementers of evaluation also reflected on the most important gaps and limitations with existing policy impact evaluations.

These were found to be limitations with adopted evaluation methodologies and designs, lack of programme theory of change, inadequate evidence-based policy-making as well as public sector budgetary constraints.

The utility value of past evaluations to policy-makers was also interrogated. It was found that these had positive contributions and benefits for policy-makings, however there were shortcomings and limitations as the evaluations did not address all the needs of policy-makers. Whilst respondents found potential value in the application of the Realist Evaluation approach, other important potential drawbacks of adopting the approach were identified. This included aspects such as: the perceived detailed and cumbersome process of Realist Evaluation; the complex nature of theory-based evaluations; the supposed lack of consensus on programme pathways to change and programme

mechanism in the general public sector which Realist Evaluation would assume to be a given fact; supposed lack of common understanding on programme input, processes and outcomes within the public sector; the vast human and financial resources required to effectively carry out a Realist Evaluation as well as feasibility challenges of possibly adopting the methods due to programme monitoring and systems that are still developing.

The evaluation approaches adopted in the public sector were found to be inappropriate due to: the types of evaluations that are currently done; the lack of attribution and causality in programme evaluation; as well as evidence of evaluation capacity issues. To test the validity of the key respondents' perceptions, it is also useful to analyse parliamentary discussions, portfolio discussions that followed the presentation of the evaluation findings; these are presented in the next chapter.

CHAPTER 7:
FINDINGS, ANALYSIS & ASSESSMENT

7.1 INTRODUCTION

This chapter presents the research findings and the emergent story based on the gathered information from the literature review, the assessment of the case studies through the lens of the Realist Evaluation Impact Assessment Framework as well as the opinions of the policy decision-makers and commissioners of impact evaluations. This enable the answering of the research question – What is the potential value of adopting a Realist Evaluation Method in programme impact evaluations in the South African public sector? Can such an approach offer new insights of what works, for whom, under what conditions and in what respects, and thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making?

In this regard, an extensive literature review on trends in programme impact evaluation including theory and practices as well as a review of the emergence of Realist Evaluation approach was undertaken. This was followed by an analysis of four impact evaluation case studies undertaken in the South African public sector, commissioned by national government departments, to assess whether the evaluation findings offered new insights of what works, for whom, under what conditions and in what respects.

Furthermore, key informant interviews were held with key informants that included policy decision-makers, commissioners and implementers of evaluations, to firstly establish what policy-makers need from evaluation findings, and secondly to capture their perception on the degree to which the completed impact evaluations met the requirements of policy-makers. The summarised results are listed in Table 7.1 to Table 7.4. This is followed by a discussion on the meanings of findings.

7.2 SUMMARY OF CASE STUDY FINDINGS

This section presents a summarised version of the data from the case studies from the applied Realist Evaluation Impact Evaluation Assessment Framework lens. Four case study analyses were undertaken. The first two cases are policy interventions on social protection, the CSG impact assessment and the youth wage subsidy evaluation. These are followed by the Grade R basic education evaluation as well as the social housing evaluation. The summarised results are presented in Table 7.1 to Table 7.4 below.

Table 7.1: Summary of Case study 1: The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households

Purpose and objectives of the evaluation	Initial programme theory of change	Impact evaluation approach and research method	Utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.			Key evaluation findings	Benefits of evaluation to policy-makers
			The programme context	Programme mechanism of change	The programme outcomes		
1. How has early versus late enrolment affected the well-being and development of children? 2. How are critical life course events of adolescents affected by the extension of the CSG? 3. What conditions determine and influence access to the CSG?	1. Cash grants targeted at children directly reduce the poverty and vulnerability of children living in poor households. 2. The CSG increases consumption and enables poor households and carers to participate in productive economic activity (e.g. to look for work).	Quasi-experimental design matching methods to establish attribution of impacts to early versus late enrolment. Participants in the study were separated into 'treatment' and 'control' groups based on whether or not they were receiving the grant and the duration of receipt.	1.Household characteristics. 2.Implementation actors and agents.	Mechanisms of the intervention in terms of how and why it worked were missing and as such, there is no evidence of collected data on appropriate mechanisms.	The first hypothesis that the cash grants targeted at children directly reduce the poverty and vulnerability of children living in poor households is partly confirmed. Infants and adolescents in poor households do not gain as much benefit as their enrolments are paltry.	The key findings focused on accessibility of the CSG to children, impacts of the CSG on young children's livelihoods, growth and human capital development as well as receipt of the CSG in adolescent households and impact of the CSG on adolescent's life outcomes.	**CSG Contributions** • The evaluation served to inform the NEPF (Commissioner #2 Social Development Evaluations, 2016). •The evaluation had a positive impact on children's lives (Commissioner #1 Social Development Evaluations, 2016). •The evaluation identified which groups were not benefiting, which served to illustrate the importance of early exposure to avoid later risky behaviour (Commissioner #1 Social Development Evaluations, 2016).

	3. The CSG addresses the underlying causes of poverty, by enabling poor households to invest in physical, social and human capital assets (education, health, nutrition), that can generate future streams of income. 4. Receipt of the CSG reduces risky behaviour by adolescents, such as transactional sex, alcohol consumption and substance abuse. 5. Specific features of the CSG (including that it is unconditional, that it targets caregivers, that it is delivered periodically and predictably, and that transaction costs are relatively low) all ensure that the overall net effectiveness of the programme social and economic impact is maximised.				The hypothesis that specific features of the CSG (including that it is unconditional, that it targets caregivers, that it is delivered periodically and predictably, and that transaction costs are relatively low) all ensure that the overall net effectiveness of the programme social and economic impact is maximised has not been supported by the reported outcomes.		• The evaluation served to strengthen the ECD policy and informed the development of a more comprehensive ECD policy (Policy Expert Social Development Evaluations, 2016). •The evaluation informed policy option regarding universalising the CSG (Commissioner #2 Social Development Evaluations, 2016). **CSG Shortcomings** •The evaluation had methodological inadequacies such as the absence of a proper control group as randomised controlled trials were not appropriate in this context (Commissioner #1 and Commissioner #2 Social Development Evaluations, 2016). •The evaluation did not address all the needs of policy-makers. Appendix F lists questions that were not clear following the evaluation •The evaluation was very costly (Commissioner #2 Social Development Evaluations, 2016). •The expertise to undertake the evaluation was outside government (Commissioner #1 Social Development Evaluations, 2016).

Table 7.2: Summary of Case study 2: A youth wage subsidy experiment for South Africa

Purpose and objectives of the evaluation	Initial programme theory of change	Impact evaluation approach and research method	Utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.			Key evaluation findings	Benefits of evaluation to policy-makers
			The programme context	Programme mechanism of change	The programme outcomes		
1. Are those with a wage subsidy more likely to be in employment as a result of the allocation of the voucher? 2. If yes, what are the mechanisms through which this effect works? 3. Do voucher holders have different types of jobs from those of non-voucher holders? 4. Does a voucher's effect persist after it has lapsed? Are there any discernible differences in the employment probabilities between voucher holders and non-voucher holders two years after	1. A school-leaver is allocated a voucher that enables any firm (subject to the firm being registered for tax and paying unemployment insurance) that decides to employ this worker to claim back a portion of the wage that the firm pays to the worker. 2. This young person searches for a job through the channels that are available to them, including their networks, formal application procedures, and informal methods such as approaching firms directly. 3. The firm chooses to experiment with an additional worker who is unable to signal their productivity, knowing that the cost of employing this worker is reduced by	Random controlled trial matching methodology where a randomly selected sample of 4 009 matched pairs of young people were randomly allocated to either the treatment or control group based on specified criteria.	1. Youth demographics, their household characteristics, educational qualifications, and previous and current labour market experiences. 2. Potential employer characteristics.	Mechanisms of the intervention in terms of how and why it worked were missing and as such, there is no evidence of collected data on appropriate mechanisms.	The initial proposition inferring that unemployed youth in possession of the voucher will use them was not proven since the case data indicated zero impact in terms of the labour force participation rate, comparing those with the voucher and those without. The theory of change proposed, "the firm chooses to experiment with an additional worker who is unable to signal their productivity, knowing that the cost of employing this worker is reduced by the	•It was found that one year after allocation, young people with the voucher were seven percentage points more likely to be in wage employment than those without the voucher. •Very few of the firms which hired young people with wage subsidy vouchers chose to use these vouchers.	**Youth wage subsidy Contributions** •The evaluation demonstrated that getting youth into the labour market early can have a positive impact on their likelihood of employment and length of employment (Levinhson et al., 2014: iv). •The evaluation informed policy choices as in 2014, the wage subsidy was implemented through the tax system as a tax incentive policy (Policy Expert Economic Cluster Evaluations, 2016). Government announced that the proposed youth employment incentive would be run through the South African Revenue Service (SARS) as a tax incentive to firms that hire young, inexperienced workers. The estimated cost to government was approximately R5 billion over three years through foregone tax revenue. The actual amount was dependent on firm uptake and job creation

voucher allocation? 5. How do firms respond to the voucher? Can their reactions inform the debate around the implementation of the wage subsidy?	the amount of the subsidy (less the administrative cost of claiming the subsidy). Through this employment, the worker gains skills and references that increase their productivity and ability to signal this productivity, which raises their income and the likelihood of being retained in employment. 5. The firm not only increases the productivity of its workforce, but also raises the productivity and reduces the uncertainty associated with the available pool of young workers.				amount of the subsidy." The analysis reveals that few firms actually utilised the subsidy and subsidy had no impact on hiring decision.		as result of the subsidy (Parliamentary Monitoring Group, 2012b). **Youth wage shortcomings** •The evaluation did not address all the needs of policy-makers. Appendix G lists questions that were not clear following the evaluation

Table 7.3: Summary of Case study 3: The impact of the introduction of Grade R on learning outcomes in South Africa

Purpose and objectives of the evaluation	Initial programme theory of change	Impact evaluation approach and research method	Utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.			Key evaluation findings	Benefits of evaluation to policy-makers
			The programme context	Programme mechanism of change	The programme outcomes		
"…whether it is possible to discern a causal effect of the introduction of Grade R on subsequent learning outcomes in school, and in particular the impact on children from disadvantaged home backgrounds."	The theory of change was not found to be explicit. Nor was a hypothesis made about what might work, for whom and in what circumstances for this particular intervention.	The study applied a quasi-experimental approach.	1. School characteristics. 2. Learner characteristics. 3. International and local early childhood development.	Mechanisms of the intervention in terms of how and why it worked were missing and as such, there is no evidence of collected data on appropriate mechanisms.	The study found that having attended Grade R, learners, from Grade 1 to Grade 6, on average improved their mathematics score by 2.5% and language proficiency by 10.2%. However, this may only applicable to well resourced, high quintile and high performing provinces. On poorer, low quintile schools, there is virtually zero impact.	The impact of Grade R in South Africa is small and there is virtually no measurable impact for the poorest three school quintiles, while there are some impacts for the higher quintile schools. Thus, instead of reducing inequalities, Grade R further extends the advantage of more affluent schools.	**Grade R contributions** •The evaluation demonstrated a plausible estimate of causal impact on those that attended Grade R and its effect on learning outcomes in later life (Sector Expert Education Evaluations, 2016). •The evaluation informed policy options regarding quality of teaching of Grade R, the quality of learning outcomes and the allocation of resources such as the remuneration of Grade R teachers (Commissioner Education Evaluations, 2016). •The evaluation highlighted some inconsistencies in the provision of Grade R and provided limited understanding of what was working and not working (Sector Expert Education Evaluations, 2016).

							Grade R shortcomings •The evaluations had limitations as it was a desktop review and issues of external validity, generalisability and legitimacy were raised “no fieldwork was done and Grade R classrooms were not visited” (Sector Expert Education Evaluations, 2016). •The evaluation did not address all the needs of policy-makers. Appendix I lists questions that were not clear following the evaluation •The design and methodology of the evaluation posed limitations as critical factors in early childhood education were not part of the evaluation (Policy Expert Social Development Evaluations, 2016).

Table 7.4: Summary of Case study 4: An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa

Purpose and objectives of the evaluation	Initial programme theory of change	Impact evaluation approach and research method	Utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.			Key evaluation findings	Benefits of evaluation to policy-makers
			The programme context	Programme mechanism of change	The programme outcomes		
To rigorously measure the impact of the programme on the welfare of local communities across a broad range of indicators.	Case study analysis did not find a hypothesis or illustrate a clear and explicit theory of change showing causal linkages between informal settlement upgrading and the life circumstance of beneficiaries and what might work for whom and in what circumstance for this particular intervention.	Quasi experimental approach to determine the causal impact between relocation from an informal settlement, in situ upgrading of an informal settlement and partial upgrade of an informal settlement against a number of well-being indicators.	Household demographics and characteristics.	Mechanisms of the intervention in terms of how and why it worked were missing and as such, there is no evidence of collected data on appropriate mechanisms.	The result found negative correlation between the formalisation of human settlements and employment levels. Given that that the Breaking New Ground Strategy has a policy objective of improving employment outcomes, the observed outcome did not support this objective.	The results show strong impacts in household demographics, asset accumulation, social interactions, satisfaction levels, household upgrading, crime rates, health and unemployment.	**Upgrading of informal settlements(UISP) contributions** •The evaluation highlighted the importance of the right combination between housing financial structuring and human settlement planning instruments which when implemented effectively can affect pathways to change (Sector Expert Human Settlements Evaluations, 2016). •The evaluation provided a general guide to debate effective methods of informal settlement upgrading (RSA, 2011:5) **Upgrading of informal settlements(UISP) shortcomings** •The evaluation did not address all the needs of policy-makers. Appendix J lists questions that were not clear following the evaluation

7.3 DETAILED CASE STUDY MICRO-ANALYSES

The research questions of this study entailed finding out:

What is the potential value of adopting a Realist Evaluation Method in programme impact evaluations in the South African public sector? Can such an approach offer new insights of what works, for whom, under what conditions and in what respects, thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making?

The literature review emphasised the value of the Realist Evaluation approach. It was found that Realist Evaluation is relevant from formative to summative programme evaluation. The strengths of Realist Evaluation come from its methodological rigour as well as its theory-based underpinnings that support better interrogation of programme impact evaluation and informs systematic analysis throughout the policy cycle. The value of the approach is embedded in its analysis of programme mechanism as well as its investigation of the programme context, as differing contextual conditions in programme design and implementation have an influence on the observed outcome patterns

Since Realist Evaluation interrogates the programme theory and asks 'what works, for whom, in what context and in what respects', this methodological approach can be applied towards impact evaluation programmes in the public sector that are implemented in large, complex, multi-faceted social environments with little or no understanding of causal mechanism. Additionally, a programme that is implemented in a different context resulting in different outcomes, even though it was implemented in the same way, could benefit from being evaluated applying this approach.

Additionally, Realist Evaluation can have useful applications in various circumstances. First, in contexts that require gaining knowledge and insights about the workings of a programme, secondly, where a programme is being implemented in a new context with no previous evidence of how it might work. Thirdly, in situations where a programme is being replicated in another context different from the previous implementation. Finally, in instances where outcome patterns are contradictory from prior implementations and the application of Realist Evaluation may serve to ascertain and provide empirical evidence of how the programme works and who can most benefit from it.

There was consensus from respondents in the interviews that the Realist Evaluation approach can potentially add value to evaluation findings. The systematic and methodical nature of the approach as well as its interrogation of programme context were seen as the two primary value-adding factors of the approach. The approach was regarded as valuable given its theory-based foundations.

The Pawson and Tilley's Realist Evaluation Method of programme evaluation is adopted as a useful tool for assessing the potential contribution of adopting a RE approach. It is being advanced as a theoretical tool of assessing the robustness of impact evaluation that are implemented in the South African public sector in terms of whether such evaluations are meaningful, valid and useful to policy-makers.

The framework reviews the purpose and objective of the evaluation, as this should be clear and coherent and provide a rationale for undertaking the evaluation. This is followed by the specification of the initial programme theory of change which specifies clearly defined pathways that the programme assumes in order to achieve the envisaged change. This theory is 'initial' since it is yet to be confirmed by the emergent programme outcome patterns when implemented. Following this the impact evaluation approach and research method are specified. The RE approach expects the research methods applied to be multi–methods, versatile, pluralist and varied as appropriate, informed by the optimum way of achieving the objectives and purpose of the evaluation.

Following this, the explanatory focus of the Realist Evaluation approach enables the assessment of the impact evaluation to generate evaluation findings that offer new insights of what works, for whom, under what conditions and in what respects. This is achieved by investigating the programme's context, in what is believed to be a complex social system, as well as the programme's mechanism of change contributing to the observed outcomes. Finally, the programme context, the mechanism of change and the observed outcome patterns, are constructed together or 'configured' to tell a coherent story that specify under what conditions, will change will be triggered resulting in the observed outcomes.

Integrating the data from both the case studies and the key informant interviews, this section will use the framework to summarise from Chapters 5 and 6 the degree to which the case studies responded to the elements of a Realist Evaluation approach.

In this regard the specific objectives are:

- To establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.
- To establish the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector.

7.4 UNARTICULATED PROGRAMME THEORY

This section presents findings on whether the programme theory was coherent and well defined and was used to define the scope and focus of the evaluation. The CSG impact evaluation had a well-articulated and graphically presented programme theory of change, mapping the receipt of the CSG to its ultimate impact of poverty reduction in the lives of beneficiaries as well as its final wider socio-economic impact. Similarly, the youth wage impact evaluation, delineated the programme's theory of change from when a school leaver is allocated a voucher up to level of the firm as an employer and its ultimate impact on productivity.

In contrast, the Grade R impact evaluation had no evidence of an explicit theory of change. In the same way, findings from the social housing impact evaluation, elicited no evidence of a clear and explicit theory of change showing causal linkages between informal settlement upgrading and the life circumstance of beneficiaries. Therefore, in some of the case studies there was a specified programme theory of change and in some this was not evident. This is a shortcoming substantiated by the interview findings. At least 57 percent of the respondents confirmed that there is an expectation of the evidence of a theory of change in policy evaluation. The theory of change was also highly prioritised by approximately 70 percent of the respondents as it articulates critical assumptions and the pathways of change of the programme. Furthermore, the literature review in Chapter 2 and the discussion in section 3.3 in Chapter 3 emphasised the important role of an articulated theory of change to provide understanding on how a programme achieves change and the importance of testing and refinement of theory which explains causality.

Across the studied cases, given that some of the propositions made were not proven by the observed outcome patterns and therefore weakened the initial programme theory, there was no evidence found regarding further development or revision of the programme theory of change. By way of example, the proposition that Grade R works for poor learners by providing stepping-stones to Grade 1 was not supported by the outcomes. Therefore, the theory of change of the Grade R programme necessitated further refinement and reframing towards alignment with the observed outcome patterns. Similarly, the theory of change of the youth wage subsidy impact evaluation was not supported by the outcomes. It was proposed that unemployed youth in possession of the voucher would actually use the voucher as conjectured. This was not proven by the outcomes since the case data indicated zero impact in terms of the labour force participation rate, comparing those with the voucher and those without. In addition, few firms actually utilised the subsidy and therefore the subsidy had no impact on the hiring decisions of the firms. These results possibly led to the revision of the policy before implementation. This is confirmed by information from expert interviews where it was indicated that “The findings were essentially process findings. It was found that voucher cards for the subsidy were cumbersome. The tax system was found to be more efficient (Policy Expert Economic Cluster Evaluations, 2016).

An explicit theory of change also serves to provide an anchor and focus the evaluation objectives. Each of the cases followed an experimental design rather than a theory-driven design. Therefore, the theory of change was not interrogated and the primary aim was to verify the internal validity of the programme as this was deemed to imply programme rationality. Consequently, the loop in these impact evaluations was not closed in terms of providing conclusive impact evaluation findings. In this instance, policy-makers are left without thorough answers; the programme works but there is no evidence on how it works. Pawson and Tilley (1997a:22) eloquently underscored this observation pointing out that the record of accomplishment of the experimental approach has mixed results and

has tended to “indicate programme success in some respects here but not there, and in other respects there but not here”.

This has been supported by the views and perceptions of the interviewed policy decision-makers, commissioners and implementers of evaluations. Some underscored these points as it was pointed out that. “Policy-makers set out interventions, but have not been able to assess failure or success. Randomised control trials are broad. We need the theory of change to make sense of programmes. The “why” is largely missing.” (Sector Expert Human Settlements Evaluations, 2016). “We don’t know what affects uptake of treatment. We do not know very much. We can harvest underlying qualitative aspects. This is becoming more prevalent. We try to explore the why of uptake. We don’t know what the uptake issues are” (Commissioner Education Evaluations, 2016).

Therefore, it does become apparent that the enduring inconclusive results of impact evaluations that have employed experimental methods weaken the utility value of evaluation results offered to policy-makers and offer limited in insights on how and why programmes work. The implication is that impact evaluations commissioned in the public sector provides inconclusive results of how programmes work, since programme theories of change drawing from both qualitative and quantitative methods are not effectively employed.

7.5 INADEQUATE PROGRAMME CAUSALITY

The CSG impact evaluation applied a quasi-experimental approach, where comparisons were constructed and made by matching households who received the treatment and those who did not, based on observable characteristics at the time of household application or enrolment into the programme. In the same way, the wage subsidy impact evaluation employed a RCT matching methodology where a randomly selected sample of matched pairs of young people were randomly allocated to either the treatment or control group, based on specified criteria and given wage subsidy vouchers to present to prospective employers who could claim a subsidy on employing the young person.

Similarly, the Grade R impact evaluation applied a non-experimental quantitative method in the form of regression analysis modelling to establish the impact of exposure to the Grade R programme on future learning outcomes of learners. In like manner, the social housing impact evaluation employed a quasi-experimental methodology to determine causal impact between relocation from an informal settlement, *in situ* upgrading of an informal settlement and partial upgrade of an informal settlement against a number of well-being indicators.

Consequently, case study analysis illustrated that all four-impact evaluations overwhelmingly applied experimental design methodologies utilising quantitative data research. Therefore, whilst approximately 71 percent of key informants ranked “the theory assumption of how change is supposed to happen” as the second most critical aspect of an evaluation, this is largely absent from public sector impact evaluations. This demonstrates the dominance of the application of

experimental methods and their enduring pervasiveness as the ‘gold standard’ in the impact evaluation of social programmes. This also suggest that impact evaluations of social programmes commissioned by the South African public sector will to a large extent adopt these homogeneous methods if there is not a concerted effort to change current thinking. It is contented that causal analysis of programme efficacy in government social programmes is presently inadequate.

This view is substantiated by key informant interviews who claimed that there are not enough evaluations that assess whether the intervention caused the results or the results can be attributed to the intervention. In this regard “the diagnostic approach that is being done is important but you need to start to do impact evaluations. Impact Evaluation must be done at a point to provide a clear roadmap.” (Policy Expert Social Development Evaluations, 2016). In addition, the literature review in Chapter 2 and the discussion in section 2.4 point to the shortcomings of experimental methods in this regard. The literature has emphasised that in contemporary evaluation practices causality is not articulated adequately.

The implication is that impact evaluations commissioned in the South African public sector applying experimental design methodologies and utilising quantitative data research methods are inadequate in providing causal links. Therefore, the South African public sector should aim to transcend the experimental design methodical paralysis that has gripped the evaluation field and accommodate alternative methods such as Realist Evaluations to enable better programme enlightenment and insights of what works. Policy decision-makers as indicated in the comments in Table 6.1, do concede that programme causality is missing and that the application of uniform methodological approached regardless of the evaluation question does not serve policy-makers in this regard.

7.6 LIMITED CONTEXTUAL UNDERSTANDING

Social programmes are influenced by a myriad of factors within their implementation setting. The broader context includes key actors and agents in the programme implementation chain, the interpersonal relationship between all programme key stakeholders; the institutional setting of the implementing agency as defined by its organisational culture and values that serve to enable or inhibit the effective implementation of a social programme, as well as overarching infrastructural system that supports the programme. All these factors can influence the direction of how the programme is implemented. This broader context, which affects the efficacy and efficiency of a programme, was found not adequately interrogated in order to gain rich insights of programme context. As the success of a programme is strongly dependent on its context, it was found that the broader contextual factors within a complex of social system were not accentuated. Virtually, across all examined impact evaluations, contextual analysis was limited to the immediate characteristics of the object of analysis, such as household demographics and characteristics, learner and school characteristics.

The CSG impact evaluation lucidly demonstrated how the integrity of the implementation chain needs to be preserved if the intended programme outcomes are to be achieved. Outcome patterns indicated that there are certain contentious issues in the implementation chain. Though the CSG is unconditional, certain application requirements impose conditions that ultimately exclude those in dire need of the grant. These impeding and exclusionary factors in the implementation chain included instances of unavailable documentary proof. This can include requirements for marriage certificates, which are usually not available in instances of customary marriages. Equally, in some rural and outlying areas, households may not have utility bills as proof of residential address. In addition, some citizens are undocumented and have yet to get birth certificates and the requisite identity documents. Similarly, some households in rural areas live far from health facilities and may not have the required health clinic cards. These challenges were apparent to the policy decision-makers and key informant interviews indicated that policy options are investigated to circumvent some of these issues. These include plans for "Government to universalise the grant to more than just qualify because of documentary challenges. In addition, other policy instruments are needed in addition to CSG to improve outcomes such as ECD" (Commissioner #2 Social Development Evaluations, 2016).

Another factor in the implementation chain pointed to the instance that some officials from the social grant disbursement agency, SASSA, due to limited information, might also contribute to gatekeeping through the provision of inaccurate information, which served to keep out eligible applicants. Alongside these factors, if the CSG was received at an alternative pay-point, it was further found to be eroded by service fees as well as compulsory spending requirements imposed by intermediaries such as supermarkets stores, who act as agents in processing payment to the beneficiaries. Therefore, these contextual conditions within the implementation of the CSG, affect the integrity of the implementation chain. That implementation can be potentially weakened with the risk of misinterpreting the programme, causing the miscarriage of the overarching policy aims.

On the other hand, in the Grade R impact evaluation there was no evidence that the social environment surrounding the implementation of the Grade R programme for both weaker and stronger provinces as well as weaker and stronger schools was investigated. It is not empirically known what circumstances in these conditions influenced and impacted on the Grade R programme nor what contextual factors make the Grade R programme thrive in one setting and fail in another. It is possible that the attitudes of parents and teachers, school leadership, provincial and district support offered to the school, the values and cultures of programme stakeholders all conspired to influence and impact on the observed Grade R outcomes. Therefore, interpersonal relationships between all programme key stakeholders as well as the institutional setting of the implementing agency as defined by its organisational culture and values should be interrogated as part of the broader context.

In the same manner, unemployed youth with wage subsidy vouchers, did not go out to the market and actively look for jobs given this support. No evidence was found interrogating this contextual factor in this evaluation.

It is possible that some of these reasons, which can be substantiated with empirical research, could have been influenced by social factors. The selected youth could have been unable to articulate the features and benefits of the wage subsidy. The cost associated with a job search could have been prohibitive for some. Therefore, the socio-economic context of the programme, the geographical provinces where the programme was implemented and other historical context could have conspired to enable or hinder programme success. These potential insights are unknown to policy-makers who simply know that the programme works within a particular margin of certainty. This is supported by the views of the key informants who highlighted contextual insights as some of most critical aspects that result in the most meaningful evaluations. About 57.1 percent of the respondents wanted policy evaluation that are clear on “The problems encountered in implementing the policy.” In addition, 43% of the respondents wanted policy evaluations that indicated, “Who benefited primarily from the policy.” In addition, the literature review in Chapter 3 section 3.3.5 emphasised how social programmes are influenced by the contexts.

It was therefore highlighted that the investigation of the context is important since this has an impact on programme outcomes. The implication from a policy perspective is that current programme impact evaluations in the South African public sector offers limited contextual understanding of the programme’s intersection within the broader complex social system.

7.7 BLACK BOX PHENOMENON

Whilst all cases were able to demonstrate whether the child support grant, Grade R, wage subsidy or the upgrade of a human settlement worked, all cases failed to show how and why each programme worked. The mechanism in which change occurs was glaring in its omission. A distinct programme mechanism, in terms of how the intervention causes change was not found in the evidence. The gathered data on the Grade R impact did not illustrate explicit programme mechanism. The underlying reasons leading to the achievement of learning outcomes were not specified. Therefore, it is not clear what it is about the Grade R programme that produces results, for whom, and under what conditions and as such the mechanisms of the intervention in terms of how and why it worked for some schools and why it did not work on others was missing. This view is substantiated by the key informant interviews where it was claimed that “We often don’t know if the programme works on average, e.g. school feeding programme, its impact on learner attendance, retention at school, cognitive outcomes and nutritional outcomes are unknown” (Sector Expert Education Evaluations, 2016).

Similarly, in both the youth wage subsidy and CSG impact evaluations, there was no evidence found of collected data on appropriate programme mechanisms and no empirical illustration of how each programme brought about change to achieve the observed outcomes. This view is substantiated by the key informant interviews as it was held that “Policy-makers set out interventions but have not been able to assess failure or success. The “why” is missing as M&E units do not have skills to do this.” (Sector Expert Human Settlements Evaluations, 2016). In addition, the literature review highlighted how the black box phenomenon in programme evaluation fails to provide enlightenment on how and why change occurs. The discussion in section 3.3.1 of Chapter 3 highlighted how theory-based methods serve to answer the questions that are in the ‘black box’ of experimental methods, since the causal links between the intervention and observed results are investigated so they can be understood.

The implication for policy-makers is that, whilst there is evident programme success, the mechanism of how that change came about is not explicit enough to inform either the further rollout or adaptation of a programme or other policy changes given insights of programme nuances in different settings. For these reasons, the mechanism of how programme results in change is largely shrouded and results in a ‘black box’ effect. The Realist Evaluation Method presents an approach that potentially lifts the lid of the box and provides illumination that can lighten the box towards a possible grey box and ultimately completely open it offering new insights of what works, for whom, under what conditions and in what respects.

7.8 UTILITY VALUE FOR POLICY-MAKERS

This section interrogates the utility value of the impact evaluation results in offering new insights of what works, for whom, under what conditions and in what respects, and thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making. In this regard, the research findings indicate that, whilst the evaluations provided policy-makers with some key insights and perspectives on the policy intervention, some gaps and limitations were evident. Whilst the CSG impact evaluation demonstrate the socio-economic benefits of providing social grants to poor and vulnerable children, there were some gaps and limitations. Infants and adolescents in poor households did no not enjoy the benefit of the CSG, as their admissions in the programme was low as there were few enrolled. Key informant interviews substantiate this view indicating, “It was found that the 0-2 years’ age group were not benefiting from the CSG. This highlighted the importance of the first 1000 days of a child's life. In addition, teenagers who receive CSG are less prone on risky behaviour and recipients of CSG stay in school longer” (Commissioner #1 Social Development Evaluations, 2016).

As a result of some of these gaps, policy-makers continued to ask questions about the CSG and its beneficiaries. An excerpt from some of the question posed to the Minister of Social Development subsequent to the impact evaluation is highlighted in Box 7.1 below. The nature of the first question asked seemed to suggest that policy-makers are concerned about access and equity of the CSG and its implementation. There is concern as to whether there is assurance that the CSG reaches the intended beneficiaries and that those who benefit are indeed the right beneficiaries. The full questions asked by policy-makers are displayed in Appendix F.

QUESTION 2755/2013

FOR WRITTEN REPLY

Date of publication on internal question paper: 18 October 2013

Internal question paper no: 33 2755.

Ms E More (DA) to ask the Minister of Social Development:

(1) During the re-registration for grants drive, how many people were found to be receiving a child support grant, despite the fact that they were not the primary caregiver of the child and the child was in fact living elsewhere;

(2) what did her department do in these cases to ensure that the child support grant was transferred to the actual caregiver and not simply suspended;

(3) how many child support grants were suspended during the re-registration drive?

Box 7.1: Questions to Minister of Social Development

Whilst the youth wage impact evaluation found that, those with a voucher had a seven percent likelihood of being in wage employment than those without the voucher a year after the allocation of the voucher, how it works and why, under what conditions it works, and for which type of young people it works was not clear. In this regard, the policy came under public scrutiny both from policy-makers and organised labour who contended that the policy might result in worker displacement and employer abuse of the subsidy scheme. Key informant interviews indicated that the impact evaluation influenced policy choices as policy-makers interrogated on whether to proceed implementing the wage subsidy through the voucher system or the tax system. It was found that were the evaluation to be rolled out at scale “a company in this study could not pick anyone. It had to be someone with the subsidy voucher” (Policy Expert Economic Cluster Evaluations, 2016). The tax system option was found to be more efficient and led into the current wage subsidy that is structured as a tax incentive policy.

Subsequently, further clarity was necessary about the implementation of the policy as well as funding mechanism, indicating possible shortcomings in the completed evaluation. An excerpt from some of the question posed to the Minister of Finance subsequent to the impact evaluation is highlighted in Box 7.2 below. Both the nature of the first and second questions asked seemed to suggest that

policy-makers are concerned about the cost of the youth wage subsidy on the budget and the implication for government in carrying these costs. The full questions asked by policy-makers are displayed in Appendix G.

In addition, The National Treasury recently issued a statement on the current status of the youth wage subsidy. It was indicated that the youth wage subsidy currently under implementation through the tax system is set to expire in December 2016 unless legislative amendments are effected for its continuation. Therefore, the policy will be subject to further evaluation at the behest of Parliament to investigate issues of uptake. Government. is unable to indicate whether the policy has been successful or otherwise as this will be informed by further evaluation and econometric studies (RSA, 2016). The full statement from the National Treasury is included as Appendix H.

QUESTION NUMBER: 1227

DATE OF PUBLICATION: 11 SEPTEMBER 2009

Mr M Swart (DA) to ask the Minister of Finance:

(1) (a) When was the pilot programme on the youth wage subsidy initiated, (b) how much money has already been spent on the programme in each month by (i) actual wage subsidy payout, (ii) monitoring costs and (iii) other costs, (c) how many persons are participating in the programme and (d) when will the scheme be implemented; (2) whether the National Treasury has any estimates on the (a) cost of implementing the youth wage subsidy and (b) number of persons to participate within the first year of its implementation; if not, what is the position in this regard; if so, what are the relevant details?

Box 7.2: Questions to Minister of Finance

On the other hand, the Grade R evaluation demonstrated a plausible estimate of causal impact on those that attended Grade R and its effect on learning outcomes in later life, but the evaluation highlighted some inconsistencies in the provision of the programme. Furthermore, the impact evaluation did not elicit crucial information on the mechanism of how and why it worked for some schools, districts or provinces and why it did not work for others, under what contexts and conditions it worked, and for which type of learners it worked. Key informant interviews substantiate this view indicating cognisance regarding these limitations. With regard to this it was emphasised that “The evaluation served its purpose and did provide limited understanding of what was working and not working. Additional gaps to knowledge are being supplemented with other evidence” (Sector Expert Education Evaluations,2016). Due to these shortcomings, further clarity about the Grade R programme and its implementation was requested. Key informant interviews indicated that the evaluation resulted in additional focus on ECD and the quality of teaching. “Subsequent to the

evaluation there was a focus on the qualification of practitioners, infrastructure issues in ECD as well as foundation phase issues and conditions" (Commissioner Education Evaluations, 2016).

This is consistent with subsequent questions asked by policy-makers after the evaluation. An excerpt from some of the question posed to the Minister of Basic Education subsequent to the impact evaluation is highlighted in Box 7.3 below. The parliamentary questions focused on the employment conditions of Grade R teachers and the recognition of their qualifications. The nature of the questions asked seemed to suggest that policy-makers are concerned about the policy issues that address the quality of teaching of Grade R. This possibly indicate the implication of this for the quality of learning outcomes and the allocation of resources such as the remuneration of Grade R teachers. The full questions asked by policy-makers are displayed in Appendix I.

DATE OF PUBLICATION FOR INTERNAL OUESTION PAPER: 02/06/2012

INTERNAL QUESTION PAPER 1512012

Miss AT Lovemore (DA) to ask the Minister of Basic Education:

(1) (a) Why have Grade R teachers not been incorporated in the staff establishment of provincial departments of education

The Education White Paper 5 on Early Childhood Development (2001) proposed a transition of providing grade R in the system through grants in aid by provincial departments of education to School Governing Bodies. This makes ECD teachers the employees of the governing bodies.

(b) Why have Grade R teachers not been recognised as teachers by the SA Council for Educators (SACE) and

The minimum qualification required for reaching in a grade R class is an ECD level 4 on the National Qualification Framework. The department has reached an agreement with the SA Council for Educators to allow for the conditional

(c) When is it anticipated that Grade R teachers will he incorporated in the provincial departments of education and recognised by SACE.

Box 7.3: Questions to Minister of Basic Education

Finally, the social housing impact evaluation found the upgrading of informal settlements in particular sites was successful in addressing some of the envisaged policy objectives. However, there was no evidence that the observed outcomes can be replicated in other settings, outside the studied sites, such as province-wide, other provinces or countrywide. There was also no evidence on which types of housing models would work for which type of households, in which contexts and under what conditions. This is substantiated by key informant interviews as there was an expressed need for evaluations that "can help us better disaggregate housing needs. It will be good to have tailored response to tailored sets of situations" (Sector Expert Human Settlements Evaluations, 2016).

Furthermore, according to key informant interviews, these gaps resulted in policy-makers interrogating and asking the best and optimum ways of funding social housing and spatial planning of human settlements.

Other subsequent questions posed to the Minister of Human Settlements, subsequent to the impact evaluation are reflected in Box 7.4 below. The questions asked whether there is a specified process that is followed for the informal settlement upgrade, how many informal settlements were upgraded, in which location these are situated and whether the resettled residents have access to basic services such as sanitation, running water and electricity. The nature of the questions asked seemed to suggest that policy-makers are concerned about how the resettlements are facilitated and implemented. In addition, there is concern about the location of these new communities and whether the upgrading is positively affecting the communities through access to better amenities and improved quality of life. The full questions asked by policy-makers are displayed in Appendix J.

NATIONAL ASSEMBLY

QUESTION FOR WRITTEN REPLY

QUESTION NO.: 1056

DATE OF PUBLICATION: 11 APRIL 2016

Ms T Gqada (DA) to ask the Minister of Human Settlements:

(1) Whether her department uses a standard checklist for the upgrading of informal settlements; if not, why not; if so,

(2) whether she will provide Ms T Gqada with a copy of the specified checklist;

(3) (a) how many informal settlements have been upgraded (i) in the (aa) 2012-13, (bb) 2013-14, (cc) 2014-15 and (dd) 2015-16 financial years and (ii) since 1 April 2016 as part of her department's informal settlements upgrade programmes, (b) what are the names of these settlements, (c) where is each specified settlement situated and (d) which services did each specified settlement receive during its upgrade?

Box 7.4: Questions to Minister of Human Settlements

Therefore, whilst all the impact evaluations demonstrated a measurable effect, policy-makers could not identify the key drivers of programme success or lack thereof. This was evidenced by the views of the interviewed policy decision-makers, commissioners and implementers of evaluations in the South African public sector as well as evidence of parliamentary questions asked by policy-makers subsequent to the studied evaluations.

This has implications for programme replication, policy review and programme enhancement as ultimately, the evaluation results must be incorporated towards better and improved policy-making. This view was substantiated by the key informant interviews which found this to be a limitation since

some evaluations are “not immediately embedded in policy-making adjustment, because the depth of what is required is not easily transferable to apply to policy adjustment. The translation requires both participant and player” (Sector Expert Human Settlements Evaluations, 2016).

The litmus test of an impact evaluation is its utility value. Getting evaluation research findings into the policy-making process is critical to influence policy change or adjustment. This view is further confirmed by key informant interviews where it was asserted that “Evaluations are not always used and there is no implementation of recommendations. Timing is a problem as evaluation are always 'late' and the results are not out on time. This results in evaluation becoming under-utilised” (Commissioner #1 Social Development Evaluations, 2016).

Clearly there is a need for evaluations to be embedded into the policy cycle timeously for maximum effect. However, policy-making is a highly-contested space. The complex and non-linear nature of the policy-making cycle and attempts at gauging the perfect timing for evaluations to feed into this process, results in evaluation utilisation that is essentially a hit or miss.

7.9 SUMMARY AND CONCLUSION TO CHAPTER 7

The purpose of this research was to explore the potential value of the REM as an approach in programme impact evaluation that result in evaluation findings that offer new insights of what works, for whom, under what conditions and in what respects. This supports the aim of evaluation findings that are meaningful, have utility value and aid in policy-making with a view towards enhancing the practice of programme impact evaluation in the South African public sector. The RE methodological approach was applied as a conceptual framework in the micro-analyses of selected case studies. Through the RE lens of context-mechanism-outcome configuration, the evaluation results of social programmes in the South African public sector were interrogated as to whether they provided useful findings and enabled policy-makers to identify key drivers of programme success.

The overall findings provided evidence that an initial well defined and coherent programme theory of change was largely absent in the commissioned impact evaluations. Pathways that map how change occurs in a programme are not articulated enough. Given that some of the propositions made were not proven by the observed outcome patterns, there was no evidence found regarding further development or refinement of the programme theory of change across all studied cases,

A consistent picture that emerged across all four cases was the prominence of applied experimental design methodologies as a chosen impact evaluation methodology. This implied that impact evaluations of social programmes commissioned by the South African public sector to date relies on a ‘one-size-fit-all’, homogeneous method. Consequently, it is contented that there is a missing explanatory focus on what causes change. Whilst there is evident programme success, the mechanism of how that change came about is not made explicit.

Furthermore, there was limited contextual understanding of the programmes' intersection within the broader complex social system. The implication being the offering of limited insights in terms of understanding whom a social programme will most effectively work for or not and the reasons thereof. This limited contextual understanding also influenced understanding of where a programme can most effectively be implemented or replicated. This had implications for nuanced understanding of the programme context.

In addition, impact evaluation in the South African public sector have a missing explanatory focus resulting in a 'black box' phenomenon where the key change mechanism in programmes are unknown since there is limited evidence found of appropriate programme mechanisms nor is there empirical evidence on how each programme brought about change to achieve the observed outcomes. It was also implied that programme impact evaluations as currently conducted has implications for the provision of useful findings resulting in evaluation findings that might not be meaningful and conclusive.

Finally, it seems to emerge that policy-makers are not well served by impact evaluation methodologies that apply 'one-size-fits-all' methods. Consistently across all four case studies, policy-makers followed the evaluation reports with further questions on the key drivers of programme success or lack thereof affecting programme planning and decision-making. This was evidenced by questions that were raised by policy-makers subsequent to the case study evaluations as more information and clarity was requested.

The results of this research has significant implications for the choice of impact evaluation methodologies that are employed in the South African public sector and the enhancement of quality impact evaluation content that resides in the National Evaluation System. This is aimed towards strengthening the National Evaluation System through programme impact evaluations findings that offer new insights of what works, for whom, under what conditions and in what respects, and thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making.

Ensuing, from this discussion, the following chapter provides an overview of the research and the conclusions drawn based on the research findings.

CHAPTER 8:
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

8.1 INTRODUCTION

The previous chapter presented and discussed the research findings and the emergent story based on the extensive literature review that was conducted, the assessment of the case studies through the lens of the Realist Evaluation Impact Assessment Framework as well as the perceptions and views of the policy decision-makers and commissioners of impact evaluations. This facilitated the answering of the research question regarding the potential value of the Realist Evaluation Method in programme impact evaluations in the South African public sector in offering new insights of what works, for whom, under what conditions and in what respects, and thereby resulting in evaluation findings that are meaningful, have utility value and aid in policy-making. Chapter 7 discussed the actual research findings and their implications.

Following from that discussion, this final chapter provides an overview of the research and the conclusions drawn based on the research findings. The importance and relevance of the research findings are discussed, as well as the significance of the research in contributing to the field of knowledge. Finally, recommendations for further research as well as future strengthening of the impact evaluations are suggested.

8.2 SUMMARY OF THE RESEARCH

The specific aims of the study were to explore and evaluate whether the Realist Evaluation Method (REM) offers potential value in the South African public sector as an additional method in programme impact evaluation. In this regard, the Realist Evaluation Impact Assessment Framework was adopted to assess existing commissioned South African government programme impact evaluations on their adopted methodologies, with the aim of answering the main research question, namely:

What is the potential value of adopting a Realist Evaluation Method in programme impact evaluations in the South African public sector? Can such an approach offer new insights of what works, for whom, under what conditions and in what respects, thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making?

The specific objectives were:

i) To gain an in-depth understanding of the Realist Evaluation Method through a detailed review and analysis of the related literature.

ii) To assess the current trends in research and application of Realist Evaluation methodical approach in conducting impact evaluations.

iii) To investigate the methodologies and approaches used in past programme impact evaluations in the South African public sector.

i) To establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.

ii) To establish the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector.

In this regard, Chapter 1 presented the rationale and context of this study within the backdrop of the international evolution towards institutionalisation of country-led M&E systems. These are driven by the demands for public accountability and the enhanced measurement of results and key policy decision-making in the public sector. South Africa's path in charting its course in institutionalising its own M&E system was reviewed. A key aspect of the system is the evaluation of programmes and policies of government as informed by the National Evaluation Policy Framework which also identifies impact evaluation as one of the main evaluation foci. The framework prescribes a National Evaluation System (NES) that implement and provide oversight over public sector evaluations. Appropriate methodological approaches in impact evaluations that can support evidence informed decision-making were extensively discussed.

Chapter 4 outlined the overarching structure of the research design and adopted methodology and provided details of the research strategy and adopted data collection methods including the data analysis, research ethics and limitations of the study. It was highlighted that the research had a three-pronged approach inclusive of comprehensive literature review, case study analyses and key informant interviews with policy-makers, commissioners and implementers of evaluations in the South African public sector. The ensuing discussion presents a summary of how these research questions and objectives were pursued, with the key conclusions to follow in the next section.

8.2.1 Objective 1: To gain an in-depth understanding of the REM through a detailed review and analysis of the related literature

Chapter 2 began by investigating trends in programme impact evaluation including the analysis of current theoretical positions and practice in both the global and local South African context. The review further advanced the discussion by providing the rationale for evidence-based policy-making and the wider evidence-based policy debate on the systematic review of evidence. The trend towards review and synthesis of evidence-based policy-making was explored and an argument was made on how such reviews and syntheses can be enhanced through the application of the Realist Evaluation approach. Within the wider evidence-based policy debate, Realist Evaluation has emerged as key contributor in the systematic review of evidence of 'what works, for whom, in what context and in what respects'. Emanating from that discussion, the review unpacked the various methodological schools of thought about the nature of evaluation and its objectives based on the literature analysis. The 'methods branch' of evaluation thinkers were found to have been influenced by the experimental design methods applied in natural sciences. Experimental design methods in evaluation emulate the experimental randomised control trials methods applied in natural sciences.

However, the methods branch of evaluation accommodates both experimental design methods and theory-driven methods.

This then gave rise to a discussion on the enduring paradigm of experimental design methodologies in the evaluation of social interventions, and an extensive discussion on the long-standing 'paradigm wars' and the utilisation of experimental methods as the 'gold standard' in conducting programme evaluations was discussed with a view to gain evidence of methodological gaps and limitations in terms of answering the question of why and how the treatment worked or not.

From thereon, the current trends in global programme impact evaluation as well as the current situation in the South African context were pondered with a view to find evidence of methodological shortcoming and gaps prevalent in current programme impact evaluation and emerging best practice. This resulted in a discussion of the methodological approaches that are applied in programme impact evaluations. Within that discussion, the current M&E landscape in South Africa both inside and outside government was discussed with a view of completing a synthesis of impact evaluation in this environment.

8.2.2 Objective 2: To assess the current trends in research and application of Realist Evaluation methodical approach in conducting impact evaluations

Stemming from that discussion in Chapter 2, Chapter 3 advanced the discussion by providing the current context of the methodological approaches in impact evaluations globally. Within that backdrop, the emergence of Realist Evaluation Method was presented, including the theoretical foundations of the approach, the key ideas of realist inquiry and its research application. In addition, the success of Realist Evaluation internationally and current international trends in Realist Evaluation were reviewed with the objective of understanding international developments, the lessons learnt, challenges and possible variations of the technique.

In this analysis, the location of the Realist Evaluation Method within the branch of theory-based evaluation was explored. Theory-driven methods specify and make explicit a programme's theory of change. Therefore, Realist Evaluation is a theory-driven evaluation programme that specifies a programme's theory of change. It has an explanatory focus which seeks to understand and interrogate 'what works, for whom, in what context and in what respects' in programme evaluation. It achieves this by specifying a configured context-mechanism-outcome (CMO) in its findings. This is achieved by understanding the context of programmes in what is believed to be a complex social system, programme mechanism in order to understand how change takes place (based on the assumption that stakeholders' reasoning influences how programmes work or their mechanism) and the interplay of both these factors resulting in observed outcome patterns.

Through its explanatory focus, Realist Evaluation is able to enlighten the 'black box' characteristic of experimental methods. Whilst these methods can demonstrate that an intervention works, they have limitation in offering a causal link as to how and why change occurred. Realist Evaluation through its testing of programme theory, using CMO configuration, potentially offers this missing link.

The discussion was progressed interrogating the application of the Realist Evaluation Method in the international public sector with international application examples highlighted. Internationally, Realist Evaluation is increasingly applied in public sector interventions across all policy environments and has been a favoured approach in healthcare sectors. Emanating from that background and critically evaluating the conditions under which to apply the Realist Evaluation approach, the review concluded by interrogating the suitability and limitations of Realist Evaluation with detailed information on the various practitioner and scholarly views about the method.

It was found that, whilst Realist Evaluation is increasingly applied in various policy environments, there remains poor application and misinterpretation of the method. Much of this misinterpretation has been the context-mechanism-outcome configuration. It was found that in attempting to specify this CMO configuration, some evaluators have come up with exhaustive fragmented 'catalogues' of plausible contexts, followed by other lists of mechanisms and yet other lists of outcomes as opposed to aligned CMOs that are built to offer coherent explanations. In response to this state of affairs, Realist Evaluation quality standards and criteria for validating true and accurate Realist Evaluation are being developed internationally. It was also found that, through its theory-based foundations that inform systematic analysis, Realist Evaluation is relevant from formative to summative programme evaluations throughout the policy cycle.

Whilst Realist Evaluation has its merits in terms of producing meticulous evaluation results and providing evidence on programme context and mechanism, it was found to require advanced theoretical understanding as well as research design and data analysis skills. As a result, it was found that the method is most suitable in complex interventions where gaining insights on programme mechanism and programme efficacy is a key objective. This include instances where programmes are implemented in new contexts with limited evidence on how they might work, where they are replicated in different contexts and in instances where the observed outcomes are contradictory from prior implementations. It is therefore concluded that in the South African public sector, Realist Evaluation could be applied in formative and summative complex interventions that are implemented in new contexts, replicated in different contexts or where programme outcomes are inconsistent. In the absence of these conditions, Realist Evaluation should not be undertaken as the method is both theoretically and labour intensive.

8.2.3 Objective 3: To investigate the methodologies and approaches used in past programme impact evaluations

Chapter 5 presented the micro-analyses of the selected case studies and addressed this objective. Micro-analyses of impact evaluation case studies were conducted to establish the methodologies and approaches used in past programme impact evaluations within the South African public sector. The first two cases were policy interventions on social protection, namely, *The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households* (RSA, 2012) and *A youth wage subsidy experiment for South Africa* (Levinhson et al., 2014). These were followed by a case study on policy intervention in basic education, *The impact of the introduction of Grade R on learning outcomes* (Van der Berg et al., 2013) and a case study on a social housing policy intervention, *An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa* (RSA, 2011b).

The micro-analyses applied the Realist Evaluation Impact Evaluation Assessment Framework as a conceptual assessment lens to evaluate each impact evaluation case study and test the robustness of the methods used in past impact evaluation in offering policy-makers the most useful findings. **Previously such extensive synthesis on current impact evaluation practice in South Africa has not been conducted and this contribution provides insights on the current environment and the impact on the National Evaluation System as a critical contribution made by this study.**

Chapter 7 integrates the case study analysis with the key informant interview data as well as the best practice guidelines derived from the literature review. A key finding from this analysis was that the presence of an initial clear and coherent programme theory on commissioned impact evaluations is largely absent and programme pathways that map how change occurs in a programme are not articulated.

It was also found that experimental design and methods are dominant in impact evaluations in the public sector. Therefore, programme causality and assessment of impact and attribution is largely missing. It was also further found that there was limited contextual understanding of the programmes' intersection within the broader complex social system resulting in possible limited understanding of whom a social programme will most effectively work for or not work for and the reasons thereof. This limited contextual understanding also influenced understanding of where a programme can most effectively be implemented or replicated.

In addition, impact evaluations in the South African public sector have a missing explanatory focus resulting in a 'black box' phenomenon where the key change mechanism in programmes are unknown since there is limited evidence found of appropriate programme mechanisms nor is there empirical evidence on how each programme brought about change to achieve the observed outcomes. It was also implied that programme impact evaluations as currently conducted have

implications for the provision of useful findings conducted resulting in evaluation findings that might not be meaningful and conclusive.

Consequently, these findings have implications for the choice of impact evaluation methodologies that are employed in the South African public sector. There is clear indication that the evaluation methods and designs are not always appropriate to inform the needs of policy-makers and consequently there are important gaps and limitations with existing policy impact evaluations.

8.2.4 Objective 4: To establish the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects.

Chapter 5, through cases study analyses established the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects. This objective was further addressed in Chapter 6 by establishing from policy decision-makers, commissioners and implementers of evaluations in the South African public sector the utility value of the evaluation results in offering new insights of what works, for whom, under what conditions and in what respects. Policy decision-makers prioritises key criteria that is expected from policy evaluations. These were deemed to be the minimum requirements required to at least results in evaluation findings that are meaningful and useful to policy decision-makers. First, valid and useful policy evaluations should at least, clearly indicate the outcomes and results of the evaluation. Secondly, the evaluation should provide a coherent programme theory of change that indicate how the programme results in the desired change. Thirdly, such evaluations should highlight any implementation challenges in order to improve programme design and enhance programme efficacy. In the fourth instance, the evaluation should provide assurance that the observed outcomes are equitable in reaching the targeted beneficiaries. Finally, the evaluation should fully present the policy implementation process.

The commissioners and implementers of evaluations in the public sector indicated that there are important gaps and limitations with existing policy impact evaluations. The first identified gap was the appropriateness of evaluation methodologies and designs as these were not always appropriate to inform the needs of policy-makers. This was followed by a lack of an articulated programme Theory of Change. This resulted in limited insights on programme pathways to change as a base of establishing how the programme works, in what context and under what conditions. Thirdly, there was perceived limited utilisation of evaluation evidence in policy-making which was seen as a limitation, as evaluation evidence was not effectively infused in the policy-making cycle. Finally, resource and budgetary constraints in the fiscus had a bearing on whether, and what type of evaluations are actually accomplished.

The utility value of past evaluations to policy-makers was also interrogated. It was found that these provided policy-makers with key insights and perspectives and had positive contributions and benefits for policy-makings, however there were shortcomings and limitations as the evaluations did not address all the needs of policy-makers. This was substantiated by evidence from the key informant interviews as well as further supported by subsequent parliamentary questions which were posed to the respective cabinet Ministers subsequent to the evaluations.

The nature of the questions asked seemed to suggest that policy-makers are concerned about the manner and the process of programme implementations, whether the intended beneficiaries do indeed benefit from the implemented policies and whether the various programme implementation chains protected the integrity of the envisaged policy aims. These questions indicated that policy-makers were somewhat still in the dark as to the identification of the key drivers of programme success, or lack thereof, and required further insights of programme success and failures.

Useful evaluation results offering new insights of what works, for whom, under what conditions and in what respect will specify key criteria that include aspects on “What were the outcomes / results of the policy evaluation”; The theory assumptions on how change is supposed to happen”; “The problems encountered in implementing the policy”; “Who benefited primarily from the policy and “How the policy was implemented”. Consequently, whilst policy decision-makers expect and need these aspects in policy evaluations, these are not always coherently presented in evaluations. **This demonstrates a misalignment between what policy decision-makers want and what they are actually getting. Therefore, the study has made some exploration in this direction, but it highlights the need for further research during the conceptualisation of impact evaluation studies to ensure that it responds to the needs of the policy-makers.**

8.2.5 Objective 5: To establish the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector

This objective was also addressed in Chapter 6 through establishing from the key informants the applicability of the Realist Evaluation Method as a methodological approach in conducting programme impact evaluations in the South African public sector. The analysis of the research findings in Chapter 7 also answered this objective. Chapter 7 presented the consolidated research findings from the literature review, the assessment of the case studies through the lens of the Realist Evaluation Impact Assessment Framework as well as the opinions of the policy decision-makers and commissioners of impact evaluations enabling the answering of the research question:

What is the potential value of adopting a Realist Evaluation Method in programme impact evaluations in the South African public sector? Can such an approach offer new insights of what works, for whom, under what conditions and in what respects, thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making?

For the reason that this research has established that impact evaluations conducted in the South African public sector in general lack a coherent programme theory, the specification of the programme's pathways to change or theory of change, highlights the critical contribution and potential value that can be made by the Realist Evaluation methodical approach in impact evaluations within the public sector. The analysis and results highlighted the crucial role that theory of change plays in mapping out the pathways of the envisaged change. An explicit theory of change also serves to provide an anchor and focus the evaluation objectives. Where the programmes pathways to changes are not coherent and articulated, causality cannot be effectively established and tested.

In addition, this research further established that impact evaluations of social programmes commissioned by the South African public sector to a large extent adopt the prevailing methods of experimental designs. These evaluations designs may not always be appropriate to effectively answer the evaluation questions, a point which was belaboured by the interviewed policy decision-makers, commissioners and implementers of evaluations. Consequently, it is contented that causal analysis of programme efficacy in government social programmes is inadequate.

This research further established that whilst evaluation findings from the various impact evaluation case studies, provided policy-makers with key insights and perspectives on these policy intervention, an explanatory focus that explained how programmes work, why the worked and for which beneficiaries under what circumstance and context was missing. These explanations are provided through the understanding of the programme's context, its mechanism of how it works and how it achieved its outcomes, a particular strength of the Realist Evaluation Method.

This analysis contributes to the core research question, in highlighting the potential value of the Realist Evaluation Method in addressing some of the gaps and limitations that are evident in government programme impact evaluation.

Given that the Realist Evaluation Method has a strong explanatory focus of how and why a programme works and since programme's pathways to change and programme causality are generally missing in some impact evaluations, the South African public sector can derive value from the Realist Evaluation Method in both formative and summative complex interventions that are implemented in new contexts, replicated in different contexts or where programme outcomes are inconsistent.

Therefore, this research has established that theory-based methods such as Realist Evaluation approach can result in evaluation findings that can potentially strengthen the South African National Evaluation System, and enable the attainment of evaluation findings that offer new insights of what works, for whom, under what conditions and in what respects, and thereby result in evaluation findings that are meaningful, have utility value and aid in evidence-informed policy-making.

To further strengthen impact evaluations included in the National Evaluation System, programme impact evaluations should not only focus on the attainment of outcomes, although the case study analyses, interviews and subsequent policy-maker discussions indicate that this is the inherent purpose of any impact evaluation. However, to address the shortcomings highlighted in Chapter 7, impact evaluations should in addition also offer insights of what works, for whom, under what conditions and in what respects, and thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making.

It is therefore concluded that the Realist Evaluation approach, through its strong explanatory focus of how and why a programme works and its accentuation of a programme's causal links, has the potential to contribute to evaluation findings that are meaningful, have utility value and aid in evidence-informed policy-making. However, there are potential constraints in its application and these should be well considered against the benefits that will be derived from such evaluations.

8.3 CONCLUSIONS ON THE RELEVANCE AND IMPORTANCE OF THE RESEARCH FINDINGS

From the research findings, it was concluded that programme theory is badly articulated and largely absent from commissioned impact evaluations. The absence of the theory of change that establish how the programme works, in what context and under what conditions as well as the clear demonstrations of programme impact results in limited insights. Therefore, the implication is that impact evaluations commissioned in the public sector might provide inconclusive results of how programmes work and are inadequate in delineating the programme pathways to change.

Secondly, it was also found that most commissioned impact evaluations applied experimental design methodologies that utilised quantitative data research as a chosen impact evaluation methodology. These methods have limitations in providing evidence of attribution and causality and ultimately finding out what works, for whom and why. Therefore, this lack of an explanatory focus results in a 'black box' phenomenon where the key change mechanism in programmes are unknown since there is limited evidence found of appropriate programme mechanisms nor is there empirical evidence on how each programme brought about change to achieve the observed outcomes.

Thirdly, it was found that most commissioned impact evaluations had limited contextual understanding of the programmes' intersection within the broader complex social system. Social programmes are influenced by a myriad of factors within their implementation setting. Gaining key insights and understanding of the broader context of the programme might bring to light, implementation challenges, that may not be initially apparent as they are embedded in the complex social structure and settings of the programme. Any unexpected programme outcomes might also be better understood through enhanced contextual understanding. This can then result in a revised and refined programme theory of change that can result in the intended outcomes.

Based on this evidence, it is concluded that the evaluation approaches that are applied in the public sector are sometimes not appropriate to inform the needs of policy decision-makers. These key stakeholders regard meaningful and valid policy evaluations as those that first, clearly indicate the outcomes and results of the evaluation, secondly, provide a coherent programme theory of change that indicate how the programme results in the desired change, thirdly, highlight any implementation challenges in order to improve programme design and enhance programme efficacy, in the fourth place, provide assurance that the observed outcomes are equitable in reaching the targeted beneficiaries, and finally construct the policy implementation process. In the absence of these, policy-makers are underserved and are left in the dark as to the identification of the key drivers of programme success or lack thereof which ultimately affect programme planning and decision-making.

In light of these findings and evidence presented, the research is offering a new consolidated model, the Realist Evaluation Impact Evaluation Assessment Model that could be applied to determine, from a Realist Evaluation lens, the value of an evaluation to different policy-maker's needs and can be applied to assess the limitations of other impact evaluation methods. This is graphically presented in Figure 8.1 and fully discussed below.

Whilst the Realist Evaluation, adapted from Pawson and Tilley was useful as a theoretical tool of assessing the robustness of impact evaluation that are implemented in the South African public sector, the framework had some gaps which the now advanced **Realist Evaluation Impact Evaluation Assessment Model aims to close. The model has three overarching aspects to it, namely: planning the evaluation, implementing the evaluation and reporting on the evaluation.**

8.3.1 Component 1: Planning the evaluation

Findings from the key informant interviews indicated that impact evaluations are currently not effectively implemented due to baseline data quality limitations. In addition, the theory of change is not articulated in programmes evaluations. The Pawson and Tilley framework which was applied in this study assumed that these systems and processes are in place, however the research evidence has shown that this is not the case. **Therefore, the revised Realist Evaluation Impact Assessment Model integrates the entrenchment of foundational baseline systems and processes to facilitate and streamline prospective impact evaluations. This assessment model is more applicable to the South African context and will be more useful to policy-makers.**

Essentially, the monitoring and baseline data should be strengthened in the intervening years whilst the programme is under implementation. Subsequently when the programme is due for impact evaluation the baseline and relevant monitoring data can facilitate the answering of Realist Evaluation questions.

Baseline systems and processes that should be embedded in programmes are the:

- Development and design of Monitoring and Evaluation framework and system;
- Development of SMART and useful KPIs for managing programme performance;
- Collection of relevant and useful qualitative and quantitative programme data;
- Development and refinement of the initial programme theory of change. If unkown or not explicit, the initial theory is to be reconstructed through interview information with programme designers, programme practitioners and staff, previous evaluation reports, programme documentary analysis as well as relevant library and desktop literature reviews.
- Programme stakeholder analysis; and
- Programme contextual analysis.

8.3.2 Component 2: Implementing the evaluation

The purpose and objectives of the evaluation should be clearly articulated as this will inform the evaluation approach to be adopted. The Realist Evaluation approach expects the evaluator to adopt research methods that are multi–method, versatile, pluralist and varied as appropriate, informed by the optimum way of achieving the purpose and objectives of the evaluation.

In addition, the case study macro-analyses which were further substantiated by the key informant interviews highlighted that evaluation exhibit a 'black box' phenomenon as the reason behind the observed change are unknown. Furthermore, not all evaluations articulated the programme theory of change which is important in understanding how the programme results in changes. The Pawson and Tilley framework was based on the assumption that there is sectorial consensus and agreement on programme theories of change, which is not entirely the case in the programme evaluation context of the public sector, a factor which was highlighted in the key informant interviews.

Case study micro-analyses showed that the programme context was superficially investigated and there was not much information on the broader programme context which has an influence on the programme outcomes. These aspects are embedded in the Pawson and Tilley context-mechanism-outcome configuration (CMO). The Pawson and Tilley framework, however assumed that these will be well understood and interpreted. The literature review has highlighted that this is not the case as these aspects of Realist Evaluation are often misinterpreted.

In this regard, the advanced Realist Evaluation Impact Evaluation Assessment Model simplifies the CMO configuration into specific activities guided by suggested questions. These questions serve to guide the evaluator and commissioners to decide whether Realist Evaluation is the type of evaluation approach that is required in response to the objectives of the evaluation.

i) Investigate broader programme context (Context)

Five core aspects should be investigated in order to understand the programme context. (1).The individual capacities of the key agents and actors who may or may not have the necessary enthusiasm, will and credibility of enabling the implementation of the social

programme; (2) the interpersonal relationship between all programme key stakeholders; (3) the institutional setting of the implementing agency as defined by its organisational culture and values that serve to enable or inhibit the effective implementation of a social programme; and (4) the overarching infrastructural system that supports it, such as political backing, resource allocation and positive public perception and support. Questions that can be posed in order to elicit this information include asking:

- Who are the key agents and actors impacting on the implementation of the programme?
- What are the individual capacities and capabilities of these agents and actors on programme implementation?
- What is the commitment of key agents and actors, their will and enthusiasm regarding the implementation of the programme?
- What are the interpersonal relationships between all programme key stakeholder?
- What factors in the institutional setting of the implementing agency enable or hinder the effective implementation of the programme?
- Does the programme implementation chain protect the integrity of the envisaged policy aims, and if not, why?
- What leadership and organisational culture factors within stakeholder organisations enable or hinder the effective implementation of the programme?
- Is there political backing or opposition to the policy programme?
- Is the programme adequately resourced?
- Is there an effective infrastructural system surrounding the programme?
- What is it about the context that prevented the anticipated change from initiating?
- To what extent does programme work or not work, for different groups of beneficiaries who are in different contextual environments?

ii) Investigate how and why the programme works (Mechanism)

- Where the programme was found to work, how did it work?
- Why does the programme work or not work?
- In what ways does the programme work for some and not work for others?
- Where the the programme did not work, why did it not work?
- How must the programme be delivered in order for it to work?

iii) Investigate the programme outcomes (Outcomes)

- For whom does the programme work and not work?
- What were the observed outcomes for the diverse intended sub-groups?
- What were the unexpected programme outcomes?
- What did the impact of demographic effects like culture had on the expected programme outcomes?

- For those who participated in the programme, do the observed outcomes patterns achieved vary across sub-groups?
- How did the way the the programme was implemented result in the observed outcomes?

The answers to these guidance questions, on programme context, mechanism and outcomes, should be properly constructed and structured to form interconnected relationships that provide insightful explanations of what works, for whom, under what conditions and in what respects. This will contribute towards evaluation findings that provide utility value to decision-makers

8.3.3 Component 3: Reporting on the evaluation

Commissioned impact evaluations in the public sector should at a minimum, results in evaluation findings that articulate the following aspects:

- The outcomes / results of the policy evaluation;
- The theory assumptions on how change is supposed to happen;
- The problems encountered in implementing the policy;
- The primary beneficiaries from the policy; and
- The main facets of the implementation approach.

Therefore, valid and useful policy evaluations should at minimum, clearly indicate the outcomes and results of the evaluation, provide a coherent programme theory of change that indicate how the programme results in the desired change, highlight any implementation challenges in order to improve programme design and enhance programme efficacy, provide assurance that the observed outcomes are equitable in reaching the targeted beneficiaries as well mapping out the policy implementation process. Furthermore, the evaluation reports should not merely pinpoint the changes observed as a result of the programme or policy intervention, but specifically highlight what brought these changes about. Such insights will add a strong dimension of usefulness in understanding for whom the programme might be most beneficial, in what context and under what circumstances. Therefore, evaluation reports that specify these components provide better enlightenment and result in meaningful, valid and useful evaluation findings that aid in evidence-informed policy-making.

The advanced Realist Evaluation Impact Evaluation Assessment Model is one of the key theoretical contributions of this study. While similar models have been proposed before, this is the first model built entirely from the review of impact evaluation studies in South Africa, thereby offering not only a significant contribution to improved Realist Evaluation design in the African context, but also to promote the ideals of African-rooted evaluation. This assessment model can be applied to assess the limitations of other impact evaluation methods in terms of whether such methods results in evaluation findings that offer new insights of what works, for whom, under what conditions and in what respects, and thereby result in evaluation findings that are meaningful, have utility value and aid in policy-making.

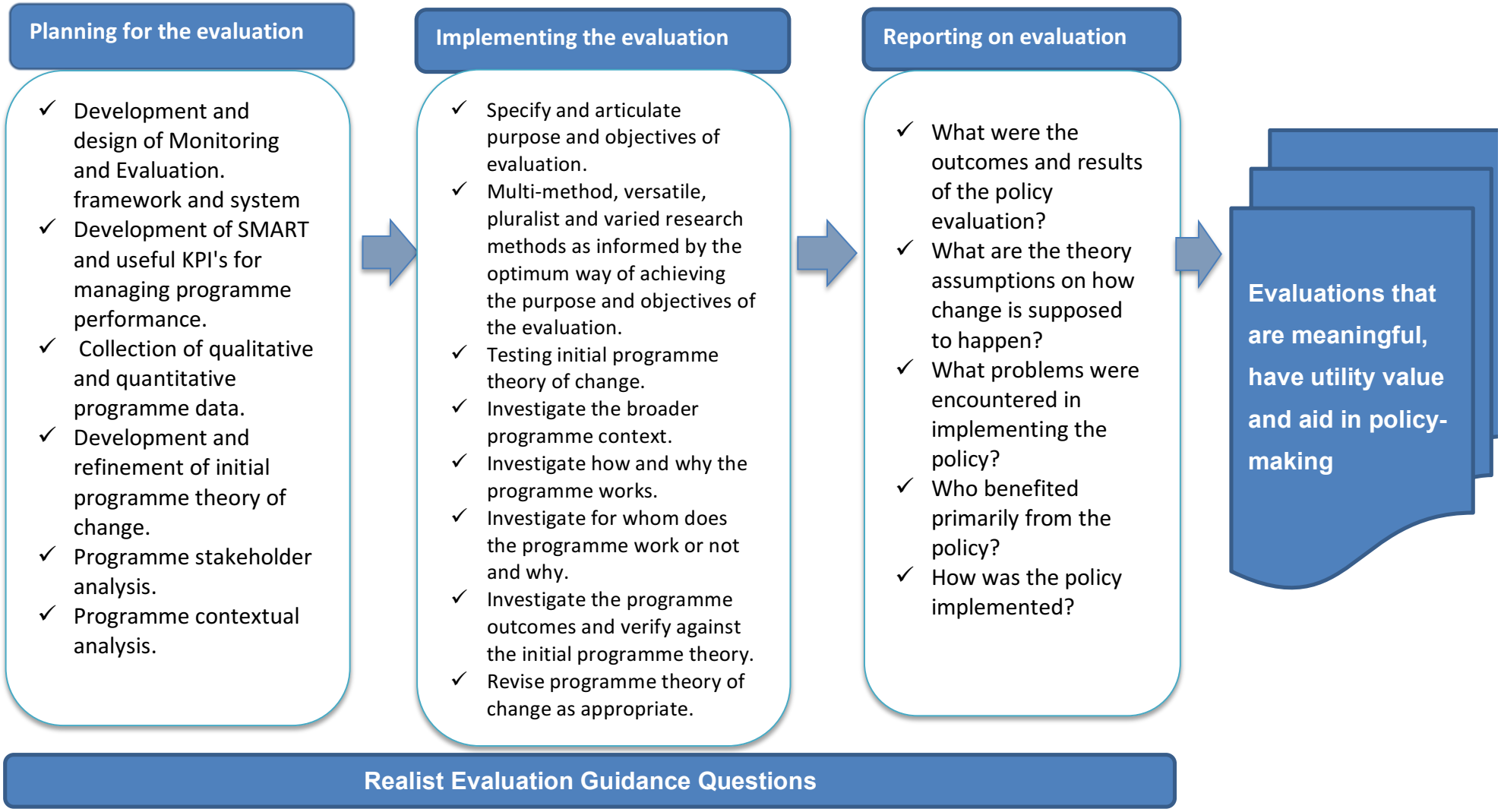


Figure 8.1 Realist Evaluation Impact Assessment Model

Source: Author.

8.4 RECOMMENDATIONS

8.4.1 Improving the National Evaluation System (NES)

From the contextual analysis in Chapter 1, it is clear that South Africa is in a critical phase of building its NES. While much progress has been made with the system in a short period of time, there is still much room for improvement, specifically in terms of impact evaluations. Currently, only a few impact evaluations have been commissioned under the NES, mostly as a result of poor implementation data and limited evaluation expertise. However, these impact evaluations answer the most critical question, 'did the intervention work' and in the future more of these evaluations will be commissioned.

8.4.2 Key evaluation questions

Impact evaluations should not only the answer the question on outcome results, but should also answer questions relevant to policy-makers, like why, for whom, under which circumstances. To do this, the designers of the evaluation should include specific questions to this effect. For example, for the four case studies analysed, the following questions may have been included in the evaluation terms of reference to pursue this aim:

i) The Child Support Grant impact evaluation

- Could the intervention work better for recipients in urban areas versus those in deep rural areas, and if so, how and why?
- Can the programme be replicated in other provinces and achieve similar or different results?
- Does the programme implementation chain protect the integrity of the envisaged policy aims, and if not, why?
- Who benefited primarily from the CSG and who did not?
- For whom does the CSG work and not work, and why? In other words within the intended beneficiaries of the CSG for which sub-group is the CSG more effective?
- Which sub-group is the CSG less effective for?
- Which sub-groups of the intended beneficiaries were reached by the CSG?
- How many of the intended beneficiaries, from which sub-groups, actively participated in the programme?
- What influenced whether the intended beneficiaries participated in the programme?
- What were the observed outcomes for the diverse intended sub-groups?
- What were the unexpected CSG programme outcomes?
- What did the impact of demographic effects like culture had on the expected programme outcomes?
- In what ways does the CSG work for some and not work for others?
- For those who participated in the CSG, do the observed outcomes patterns achieved vary across sub-groups? Why?

- To what extent does the CSG work or not work, for different groups of beneficiaries who are in different contextual environments?
- Where the CSG was found to work, how did it work? - What were the programme mechanism that led to change?
- Where the CSG did not work, why did it not work?
- For whom did the anticipated change mechanisms not activate?
- What is it about the context that prevented the anticipated change mechanisms from activating?
- How must the CSG be delivered in order for it to work?
- How did the way the CSG was implemented result in the observed outcomes?
- Why did some intended beneficiaries not benefit from the CSG?

ii) The Youth Wage Subsidy impact evaluation

- Could the youth wage subsidy work better for recipients in urban areas versus those in deep rural areas, and vice-versa, if so, how and why?
- Could the youth wage subsidy work better for other sub-groups other than African males?
- Can the programme be replicated in other provinces and achieve similar or different results?
- Does the programme implementation chain protect the integrity of the envisaged policy aims, and if not, why?
- Who benefited primarily from the youth wage subsidy and who did not?
- For whom does the youth wage subsidy work and not work, and why? In other words within the intended beneficiaries of the youth wage subsidy for which sub-group is the youth wage subsidy more effective?
- Which sub-group is the youth wage subsidy less effective for?
- Which sub-groups of the intended beneficiaries were reached by the youth wage subsidy?
- How many of the intended beneficiaries, from which sub-groups, actively participated in the programme?
- What influenced whether the intended beneficiaries participated in the programme?
- What were the observed outcomes for the diverse intended sub-groups?
- What were the unexpected programme outcomes?
- What did the impact of demographic effects like culture had on the expected programme outcomes?
- In what ways does the youth wage subsidy work for some and not work for others?
- For those who participated in the youth wage subsidy, do the observed outcomes patterns achieved vary across sub-groups? Why?
- To what extent does the youth wage subsidy work or not work, for different groups of beneficiaries who are in different contextual environments?

- Where the youth wage subsidy was found to work, how did it work? - What were the mechanism that led to change?
- Where the youth wage subsidy did not work, why did it not work?
- For whom did the anticipated change mechanisms not activate?
- What is it about the context that prevented the anticipated change mechanisms from activating?
- How must the youth wage subsidy be delivered in order for it to work?
- How did the way the youth wage subsidy was implemented result in the observed outcomes?
- Why did some intended beneficiaries not benefit from the youth wage subsidy?

iii) The Grade R impact evaluation

- Can the programme be replicated in other provinces and achieve similar or different results?
- Does the programme implementation chain protect the integrity of the envisaged policy aims, and if not, why?
- Within the intended beneficiaries of the Grade R programme, for which sub-group is the programme more effective for?
- Which sub-group is the Grade R programme less effective for and why ?
- Which sub-groups of the intended beneficiaries were reached more effecctively and less effectively by the Grade R programme?
- How many of the intended beneficiaries, from which sub-groups, actively participated in the Grade R programme?
- What influenced whether the intended beneficiaries participated in the Grade R programme?
- What were the observed outcomes for the diverse intended sub-groups?
- What were the unexpected programme outcomes?
- What did the impact of demographic effects like culture had on the expected programme outcomes?
- In what ways does the Grade R programme work for some and not work for others?
- For those who participated in the Grade R programme, do the observed outcomes patterns achieved vary across sub-groups? Why?
- To what extent does the Grade R programme work or not work, for different groups of beneficiaries who are in different contextual environments?
- Where the Grade R programme was found to work, how did it work? - What were the programme mechanism that led to change?
- Where the Grade R programme did not work, why did it not work?
- For whom did the anticipated change mechanisms not activate?
- What is it about the context that prevented the anticipated change mechanisms from activating?

- How must the Grade R programme be delivered in order for it to work effectively?
- How did the way the Grade R programme was implemented result in the observed outcomes?
- Why did some intended beneficiaries not benefit from the Grade R programme?

iv) The upgrade of informal settlements(UISP) impact evaluation

- Can the UISP programme be replicated in other provinces and achieve similar or different results?
- Does the UISP programme implementation chain protect the integrity of the envisaged policy aims, and if not, why?
- Within the intended beneficiaries of the UISP programme, for which sub-group is the programme more effective for?
- Which sub-group is the UISP programme less effective for and why ?
- Which sub-groups of the intended beneficiaries were reached more effectively and less effectively by the UISP programme?
- How many of the intended beneficiaries, from which sub-groups, actively participated in the UISP programme?
- What influenced whether the intended beneficiaries participated in the UISP programme?
- What were the observed outcomes for the diverse intended sub-groups?
- What were the unexpected programme outcomes?
- What did the impact of demographic effects like culture had on the expected programme outcomes?
- In what ways does the UISP programme work for some and not work for others?
- For those who participated in the UISP programme, do the observed outcomes patterns achieved vary across sub-groups? Why?
- To what extent does the UISP programme work or not work, for different groups of beneficiaries who are in different contextual environments?
- Where the UISP programme was found to work, how did it work? - What were the mechanism that led to change?
- Where the UISP programme did not work, why did it not work?
- For whom did the anticipated change mechanisms not activate?
- What is it about the context that prevented the anticipated change mechanisms from activating?
- How must the UISP programme be delivered in order for it to work effectively?
- How did the way the UISP programme was implemented result in the observed outcomes?
- Why did some intended beneficiaries not benefit from the UISP programme?

8.4.3 Alternative methodologies

It is important to pursue a flexible range of alternative methodologies in impact evaluation. At present, the quantitative nature of some of the commissioned impact evaluations limits the usefulness of the findings. The ineffective 'one-size-fits-all' method of addressing interventions does not serve the NES. The application of inflexible evaluation methodologies regardless of their relevance to the evaluation question limit the level of enlightenment that could be derived from such evaluation findings. Hence, the South African public sector should strive to stimulate the accommodation of alternative methodologies whose application can effectively serve the needs of policymaking.

8.4.4 Evaluation expertise

Related to the above recommendation, there is a need to enhance the evaluation expertise in the public sector, as well as in the broader evaluation community. As the NES is further strengthened, it is expected that there will be an increasing demand for impact evaluations. As a result of these factors, public sector managers will have to plan for the commissioning of programme evaluations and impact assessments in line with the prescript of the NES. Therefore, the South African public sector must provide enhanced intellectual leadership in developing evaluation expertise and skills. This will demand organic evaluation capacity building which will come from various training and skills developments initiatives. Tackling these current capacity constraints within the public sector demand a multi-stakeholder approach involving actors like the DPME, SAMEA, PSC, the National School of Government, public universities and other learning institutions. Such collaborations will foster the design of relevant monitoring and evaluation curriculum which is relevant and speaks to the needs of government.

8.4.5 Consolidation of evaluation criteria

There is a need for further engagement with policy-makers and the users of evaluation to consolidate a list of criteria that will enhance the usefulness and value of impact evaluations, as this research reveals a misalignment between the questions raised by policy-makers and the content of the evaluation reports. Actors such as DPME, a primary user of evaluations, should lead the process in collaboration with other stakeholders towards improving the quality and usefulness of evaluation emanating from the NES.

8.4.6 The potential value of Realist Evaluation approach

The benefits that can be derived from a Realist Evaluation Method are: first, its findings are meticulous and in-depth due to the systematic and theoretical rigour of the approach. Such findings can support the tailoring of defined interventions, targeting specific beneficiaries. Secondly, the approach interrogates the programme context as well as how, why and for whom the programme works. Therefore, it supports programme evaluation that is evidence-based, demonstrating contextual conditions required for programme success. Thirdly, the benefit of the method is that it supports evidence informed policy-making. Within the wider evidence-based policy debate, Realist

Evaluation has emerged as key contributor in the systematic review of evidence of 'what works, for whom, in what context and in what respects'. This strong base of evidence can be used to support or challenge policy and programme claims of what it is that works and provide a menu of policy options in its findings.

However, whilst the method produces meticulous, and in-depth findings, the application of this method should be well considered to avoid the following potential constraints. Firstly, the approach requires intensive investment in time and resources as it involves rigorous testing of programme theory as the evaluator must thoroughly investigate programme context, mechanism and outcomes. Therefore, the approach is far from a tick box exercise. Secondly, the approach requires advanced understanding of programme theory and requires good research skills as the evaluator must make decisions on how to design the research and choose what data to research. In this regard it can be intellectually challenging.

The interviewed respondents also identified potential drawbacks of adopting a Realist Evaluation approach. Initially, it was emphasised that, as a theory-based evaluation approach, Realist Evaluation will require an understanding of the detailed methodical process by evaluators and commissioners alike. Furthermore, as the approach has a strong focus on programme theory, there will be a need for consensus amongst all stakeholders regarding what the programme pathways to change are within the specific policy sectors. Additionally, the practicalities of factoring enough lead time to effectively execute the evaluation as well as the availability of adequate financial resources to fund such evaluation will have to be well considered.

Therefore, a Realist Evaluation design cannot be a regular part of all conducted evaluations, since the method is quite thorough. Some evaluations do not require such a level of depth and rigour to answer the evaluation questions and may require less probing strategies. Consequently, the application of a Realist Evaluation should be considered in instances where there is a desire to learn and gain knowledge as the method produces meticulous and thorough findings. Therefore, the costs of a Realist Evaluation should be balanced against the benefits that will be derived from such an evaluation.

It therefore seems that a Realist Evaluation design may be most useful in the following cases: *First*, where evaluation questions are asked that seek to find knowledge and insight about the workings of a programme. *Second*, where a programme is being implemented in a new context with no previous evidence of how it might work. *Third,* where a programme is being adapted in a different context. *Fourth*, in instances where outcome patterns are contradicting prior implementations. The approach may serve to confirm and provide empirical evidence of how the programme works, why, under what circumstances, and who can most benefit from it. *Fifth*, it can be applied towards impact evaluations that are implemented in large, complex, multi-faceted social environments with little or no understanding of causal mechanism.

To promote and ensure the uptake of the approach it is recommended that:

- An explicit reference to Realist Evaluation should be made in the evaluation TOR as an appropriate methodology;
- Stakeholders involved in the framing of the evaluations questions should be broadened to include members of parliamentary portfolio committees who are mandated to provide oversight and policy directions on government performance;
- The engagement of local and international thought leadership on Realist Evaluation as part of ongoing M&E skills development and capacity building should be promoted.
- Programme ToC should be required in all project designs as a matter of routine, before approval of the project. This will facilitate the impact evaluation and would also allow for testing the validity of the ToC conceptual and action dimensions during the Realist Evaluation.

8.5 LIMITATIONS OF THE STUDY

8.5.1 Case study selection

Due to the limited scope of impact evaluations drawn from the public sector, the sample is justifiably small. Since the population is small a large sample is not drawn. Due to the small sample population, these four cases are deemed to bring relevant insights and perspectives for the objective of establishing the methodologies and approaches used in past programme impact evaluations. Whilst a thorough search of impact evaluations completed for the South African public sector were conducted, informed by the resources available to this researcher, it is possible that besides the identified completed impact evaluations selected for this study, there are other completed impact evaluations commissioned by the state that could have been missed. These could be completed impact evaluations outside the NES repository of the DPME, those that are not publicly available, as well as completed impact evaluations commissioned by the state and conducted by other parties other than the identified international development institutions.

8.5.2 Key informant interviews

Gaining access to key respondents such as the relevant ministerial advisor who would have policy insights on some of the policies informing the impact evaluations in this study as well as some of the DPME outcome facilitators who monitor the implementation of policy outcomes was difficult. Through persistence following up, in the final stage there were a total of seven respondents who granted access and were interviewed. Whilst this is a limited number, these were policy experts in the relevant areas of this study who commissioned evaluations and drafted policies. Three were experts in basic education policy evaluations, two were experts in social development and had commissioned the relevant evaluation that is part of this study, one was a DPME government outcome facilitator who provided broad perspective and insights on the role of impact of evaluations in government in informing progress on policy outcomes as well as human settlements policy evaluations. The last respondent was an expert who was involved in youth employment policy

evaluations who provide key insights on the relevant evaluation that is part of this study as well as current perspectives on the implementation of youth employment policies.

Whilst this limited number of interviews possibly provided fewer perspectives and insights, these respondents were able to offer input covering all four case studies. Some of the gaps in policy insights were mitigated by a search for specific parliamentary questions, which were asked by policy-makers in relation to the selected impact evaluations in this study. The relevant parliamentary questions and answers provided are included in Appendices F, G, I and J.

8.6 SUMMARY AND CONCLUSION TO CHAPTER 8

This chapter provided a summary of the research, drew conclusions on the relevance and importance of the research findings, their implication within the context of the defined research problem and stated research objectives identified in Chapter 1. The significance of the work in contributing to the field of knowledge was discussed including the advancement of a model, more applicable to the South African context, that facilitate and streamline prospective impact evaluations. Recommendations for the field as informed by the research finding were made. Finally, the limitations of the study were discussed.

REFERENCES

Abrahams, M.A. 2003. *Theory-based evaluation of community development: A South African case study*. Unpublished PhD dissertation. Bellville: Stellenbosch University.

Abrahams, M.A. 2015. A review of the growth of monitoring and evaluation in South Africa: Monitoring and evaluation as a profession, an industry and a governance tool', *African Evaluation Journal,* **3**(1), 1-8.

Alkin, M.C. & Christie, A.C. 2004. An Evaluation Theory Tree. In Alkin, M.C., *Evaluation Roots: Tracing theorists' views and influences*. University of California, LA: Sage Publications.

Alkin, M.C. (ed.) 2004. *Evaluation Roots*. Thousand Oaks, CA: Sage Publications. [Online] Available: doe: http://dx.doi.org.ez.sun.ac.za/10.4135/9781412984157 Accessed: 11 May 2015.

Alkin, M.C. (ed.) 2013. *Evaluation Roots: A wider perspective of theorists' views and influences*. 2nd edition. Thousand Oaks, CA: Sage.

Alkin, M.C. 1990. *Debates on Evaluation*. Newbury Park, CA: Sage Publications.

Astbury, B. 2013. Some reflections on Pawson's Science of Evaluation: A Realist Manifesto. *Evaluation,* **19**(4) 383-401.

Archer, M., Bhaskar, R., Collier, A., Lawson, T. & Norrie, A. (eds.). 1998. *Critical Realism: Essential Readings*. London and New York: Routledge.

Auriacombe, C.J. 2013. In search of an analytical evaluation framework to meet the needs of governance. *Journal of Public Administration*, **48**(4.1), 715-729.

Bamberger, M. 2008. Enhancing the utilization of evaluations for evidence-based policy-making. In Segone, M., *Bridging the gap: The role of monitoring and evaluation in evidence-based policy-making*. Evaluation Working Paper #12. New York: UNICEF, 120-142.

Bamberger, M., Vijayendra, R. & Woolcock, M. 2010. *Using mixed methods in monitoring and evaluation: Experiences from international development*. World Bank Policy Research Working Paper 5245, March.

Bellagio Report. 2013. *The Bellagio Report*. African Thought Leaders Forum on Evaluation and Development: Expanding Leadership in Africa. The Bellagio Centre. [Online] Available: www.clear-aa.co.za/wp-content/uploads/2013/09/Bellagio-Report-Vs-Apr-14.pdf. Accessed: 11 June 2017

Berk, R. 2011. Evidence-based versus junk-based evaluation research: some lessons from 35 years of the Evaluation Review. *Evaluation Review,* **35**(3), 191-203

Best, A., Greenhalgh, T., Lewis, S., Saul, J.E., Carroll, S. & Bitz, J. 2012. Large-system transformation in Health Care: A realist review. *The Milbank Quarterly*, **90**(3), 421-456.

Betts, J. 2013. Aid effectiveness and governance reforms: Applying realist principles to a complex synthesis across varied cases. *Evaluation,* **19**(3), 249-268.

Bhaskar, R. 2008. *A Realist Theory of Science.* London and New York: Routledge.

Birckmayer, J. & Weiss, C., 2000. Theory-based evaluation in practice: what do we learn? *Evaluation Review,* **24**, 407-431.

Blamey, A. & Mackenzie, M. 2007. Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation,* **13**(4), 439-455.

Boruch, R, 2005. Preface: Better evaluation for evidence-based policy: Place randomized trials in education, criminology, welfare and health. *Annals of the American Academy of Political and Social Science*, **599**, 6-18.

Boruch, R., McSweeney, A. & Soderstrom, E. 1978. Randomized field experiments for program planning, development and evaluation. *Evaluation Quarterly,* **2**(4), 655-695.

Brown, C. 2014. Advancing policy-makers' expertise in evidence-use: A new approach to enhancing the role research can have in aiding educational policy development. *Journal of Educational Change,* **15**(1),19-36.

Cameron, D.B., Mishra, A. & Brown, A.N. 2016.The growth of impact evaluation for international development: How much have we learned? *Journal of Development Effectiveness*, **8**(1), 1-21.

Campbell, D. & Stanley, J. 1966. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Campbell, D. 1957. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, **54**(4), July, 297-312.

Carden, F., Bamberger, M. & Rugh, J. 2009. *Alternatives to the Conventional Counterfactual: Follow-up note on Session 713 Think Tank*. American Evaluation Association, Orlando. [Online] Available http://www.realworldevaluation.org/uploads/Alternative_approaches_to_the_counterfactual_AEA_09.doc. Accessed: 13 June 2017.

Carden, F. & Alkin M.C., 2012. Evaluation roots: An international perspective. *Journal of Multi-Disciplinary Evaluation*, **8**(17), 102 -118

Center for Global Development. 2006. *When will we ever learn? Improving Lives Through Impact Evaluations*. Report of the Evaluation Gap Working Group. New York. [Online] Available: http://www.cgdev.org/content/publications/detail/7973. Accessed: 18 June 2017

Chelimsky, E. & Shadish, W.R. (eds.). 1997. *Evaluation for the 21st century. A handbook*. Thousand Oaks, CA: Sage Publications.

Chelimsky, E. 1997. The coming transformations in evaluation. In Chelimsky, E. & Shadish, W.R. (eds.). *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: Sage Publications, 1-27. doi: http://dx.doi.org.ez.sun.ac.za/10.4135/9781483348896.n1

Chen, H. T. & Rossi, P. 1983. Evaluating with sense: The theory-driven approach. *Evaluation Review*, **7**(3), 283-302.

Chen, H.T. 1989. The conceptual framework of the theory driven perspective. *Evaluation and Programme Planning*, **12**(4), 391-396.

Chen, H.T. 1990. *Theory-driven evaluations*. Thousand Oaks, CA: Sage Publications.

Chen, H.T. 1994. Current trends and future directions in programme evaluation. *American Journal of Evaluation*, **15**(3), October, 229-238.

Chen, H.T. 2005. Theory-driven evaluation. In Mathison, S. (ed.), *Encyclopaedia of Evaluation*. Thousand Oaks, CA: Sage Publications, 416-421. [Online] Available: doi: http://dx.doi.org.ez.sun.ac.za/10.4135/9781412950558.n542 Accessed: 11 May 2015.

Chen, H.T. 2005. *Practical Programme Evaluation.* Thousand Oaks, CA: Sage Publications.

Chen, H.T. 2006. A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, **13**(1), Spring, 75-83.

Chen, H.T. 2010. The bottom-up approach to integrative validity: A new perspective for programme evaluation. *Evaluation and Programme Planning*, **33**, 205-214.

Chilisa, B.& Malunga, C. 2012*. Made in Africa Evaluation: Uncovering African Roots in Evaluation Theory and Practice,* 32-38. In Bellagio Report. 2013. *The Bellagio Report*. African Thought Leaders Forum on Evaluation and Development: Expanding Leadership in Africa. The Bellagio Centre. [Online] Available: www.clear-aa.co.za/wp-content/uploads/2013/09/Bellagio-Report-Vs-Apr-14.pdf. Accessed: 11 June 2017

Chilisa, B., Major, T.E., Gaotlhobogwe, M. & Mokgolodi, H., 2016. Decolonizing and indigenizing evaluation practice in Africa: toward African relational evaluation approaches. *Canadian Journal of Program Evaluation*, ***30***(3),313-328.

Chilisa, B., Major, T.E., & Khudu-Petersen, K., 2017. Community engagement with a postcolonial, African-based relational paradigm. *Qualitative Research*, ***17***(3),326-339.

Clarke, A. & Dawson, R. 1999. *Quantity and Quality: Paradigmatic Choices in Evaluation Methodology in Evaluation Research: An introduction to Principles, Methods and Practice*. Thousand Oaks, CA: Sage Publications.

Cloete, F. 2009. Evidence-based policy analysis in South Africa: Critical assessment of the emerging Government-Wide Monitoring and Evaluation System. *Journal of Public Administration,* **44**(2), 293-311.

Cloete, F. 2016. Developing an Africa-rooted programme evaluation approach. *African Journal of Public Affairs,* **9**(4), 55-70.

Cook, T.D. & Campbell, D.T. 1979. *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally

Coryn, C.L.S., Hattie, J.A., Scriven, M. & Hartmann, D.J. 2007. Models and Mechanisms for Evaluating Government-Funded Research: An International Comparison. *American Journal of Evaluation,* **28**(4), 437-457.

Coryn, C.L.S., Noakes, L.A., Westine, C.D. & Schroter, D.C. 2011. A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation,* **32**(2), 199-226.

Cronbach, L. & Associates. 1980. *Toward reform of programme evaluation: Aims, methods, and institutional arrangements*. San Francisco, CA: Jossey-Bass.

Cronbach, L. 1982. *Designing evaluations of educational and social programmes*. San Francisco, CA: Jossey-Bass.

Davies, H., Nutley, S. & Smith, P. 2000. *Introducing evidence-based policy and practice in public services. What Works?* Bristol: Policy Press.

Davis, P. 2005. The limits of realist evaluation: Surfacing and exploring assumptions in assessing the best value performance regime. *Evaluation,* **11**(3), 275-295.

Duflo, E., & Kremer, M. 2003, *Use of Randomization in the Evaluation of Development Effectiveness*. World Bank Operations Evaluation Department. World Bank, Washington, DC.

Earl, S., Carden, F. & Smutylo, T. 2001. *Outcome Mapping: Building learning and reflection into development programs*. International Development Research Centre: Canada.

Engela, R. & Ajam, T. 2010. *Implementing a Government-wide Monitoring and Evaluation System in South Africa*. Independent Evaluation Group, ECD Working Paper Series 21. World Bank, Washington, DC.

Fetterman, D.M. 1994. Steps of Empowerment valuation: From California to Cape Town. *Evaluation and Programme Planning*, **17**(3), 305-313.

Floyd, M., Pilling, D., Garner, K. & Barrett, P. 2004. Vocational rehabilitation: What works and in what circumstances. *International Journal of Rehabilitation Research*, **27**(2), 99-103.

Fournier, D. 2005. Evaluation. In Mathison, S. (ed.), *Encyclopaedia of Evaluation*. Thousand Oaks, CA: Sage Publications, 140-141. [Online] Available: doi: http://dx.doi.org.ez.sun.ac.za/10.4135/9781412950558.n179. Accessed: 11 May 2015.

Gaarder, M. & Briceño, B. 2010. Institutionalisation of government evaluation: Balancing trade-offs. *International Initiative for Impact Evaluation*. Working Paper 8.

Gertler, P.J., Martinez, S., Premand P., Rawlings, L.B. & Vermeersch, C.M.J. 2011. *Impact Evaluation in Practice.* World Bank, Washington, DC. [Online] Available http://www.worldbank.org/ieinpractice. Accessed: 18 June 2017.

Greenhalgh, T., Humphrey, C., Hughes., MacFarlane., Butler, C. & Pawson, R. 2009. How do you modernize a health service? a realist evaluation of whole-scale transformation in London. *Milbank Quarterly,* **87**(2), 391-416.

Greenhalgh, T., Wong, G., Jagosh, J., Greenhalgh, J., Manzano, A., Westhorp, G. & Pawson, R. 2015. *Protocol – the RAMESES II study: Developing guidance and reporting standards for Realist Evaluation*: BMJ Open **5**(e008567) [Online] Available: doi:10.1136/bmjopen-2015- 008567 Accessed: 7 January 2016.

Greenhalgh, T., Wong, G., Westhorp, G. & Pawson, R. 2011. Protocol – Realist and Meta-Narrative Evidence Synthesis: Evolving Standards (RAMESES). *BMC Medical Research Methodology,* **11**(115).

Hair Jr., J.F., Celsi, M.W., Money, A.H., Samouel, P. & Page, M.J.2011. *Essentials of Business Research Methods*. 2nd edition. M.E Sharpe. Armonk: New York

Hawkes, S., Zaheer, H., Tawil, O., O'Dwyer, M. & Buse, K. 2012. Managing research evidence to inform action: Influencing HIV policy to protect marginalised populations in Pakistan. *Global Public Health: An International Journal for Research, Policy and Practice*, **7**(5), 482-494.

Haynes. L., Service, O., Goldacre, B. & Torgerson, D. 2012. *Test, Learn, adapt: Developing Pubic Policy with Randomised Controlled Trials*. London: Cabinet Office. [Online] Available: www.cabinetoffice.gov.uk. Accessed 1 November 2016.

Hayton, K. 2015. Evaluating the evidence – a move to more Realistic Evaluation: A case study of Regional Selective Assistance in Scotland. *Evaluation,* **21**(2), 248-262.

Head, B.W.2016. Toward more “evidence-informed” policy making? *Public Administration Review,* **76**(3), 472-484.

Henry, G. 2005. Realist Evaluation. In Mathison, S. (ed.), *Encyclopaedia of Evaluation*. Thousand Oaks, CA: Sage Publications, 360-363. [Online] Available: doi: http://dx.doi.org.ez.sun.ac.za/10.4135/9781412950558.n179. Accessed: 11 May 2015.

Henry, G.T. & Mark, M.M. 2003. Toward an agenda for research on evaluation. In Christie, C.A. (ed.), The practice-theory relationship in evaluation. *New Directions for Evaluation series*, 97. San Francisco, CA: Jossey-Bass, 69-80.

House, E. 1991. Evaluation and social justice: Where are we? In McLaughlin, M.W. & Phillips, D.C. (eds.), *Evaluation and education: At quarter century,* 90th yearbook of the National Society for the Study of Education, Part II, 233-247. Chicago: University of Chicago Press.

House, E. 1993. *Professional evaluation: Social impact and political consequences*. Newbury Park, CA: Sage Publications.

House, E. 2001. Unfinished business: Causes and values. *The American Journal of Evaluation*, **22**(3), 309-315.

House, E. & Howe, K. 2000. Deliberative Democratic Evaluation. *New Directions for Evaluation*, **85**, Spring, 3-12.

Hudson, B. 2001. Book review of Davies, H., Nutley, S. & Smith, P. 2000. Introducing evidence-based policy and practice in public services. What Works? Bristol: Policy Press. *Health and Social care in community*, **9**(6), 504-511.

Johnson, R.B. 1998. Toward a theoretical model of evaluation utilization. *Evaluation and Programme Planning*, **21**(1), 93-110.

Lavis, J.N. 2009. How can we support the use of systematic reviews in policy-making? *PLoS Medicine,* **6**(11): e1000141.

Lee, N. & Lings, I.2008. *Doing Business Research: A Guide to Theory and Practice*. Thousand Oaks, CA: Sage Publications.

Levin, R.M. 2009. Transforming the public service to support the developmental state. *Journal of Public Administration*, **44**(4.1) 943-968.

Levinhson, J., Rankin, N., Roberts, G. & Schöer, V. 2014. A youth wage subsidy experiment for South Africa. *Impact Evaluation Report* (**15**) 1-74. International Initiative for Impact Evaluation. [Online] Available: http://www.3ieimpact.org/en/publications/3ie- impact-evaluation-reports/3ie-impact-evaluations. Accessed: 1 August 2016

Lincoln, Y.S. & Guba, E.G.1986. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Programme Evaluation*, **30**, 73-84**.**

Lopez-Acevedo, G., Krause, P. & Mackay, K.2012. *Building Better Policies: The Nuts and Bolts of Monitoring and Evaluation Systems.* The World Bank. Washington

Marchal, B., Dedzo, M. & Kegels, G. 2010. Turning around an ailing district hospital: a realist evaluation of strategic changes at Ho Municipal Hospital (Ghana). *BMC Public Health*, **10**, 787. [Online] Available: 10.1186/1471-2458-10-787 Accessed: 1 August 2016.

Marchal, B., Van Belle, S., Van Olmen, J., Hoerée, T. & Kegels, G. 2012. Is Realist Evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation,* **18**(2), 192-212.

Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G. & Pawson, R. 2013. Realist RCTs of complex interventions – an oxymoron. *Social Science & Medicine,* **94**, 124-128.

Mark, M.M., Henry, G.T. & Julnes, G. 1999. Toward an integrative framework for evaluation practice. *American Journal of Evaluation*, **20**(2), 177-198.

Masikane, R. 2016, email, 6 May, Refilwe.Masikane@dpme.gov.za

Mathison, S. (ed.) 2005. *Encyclopaedia of Evaluation*. Thousand Oaks, CA: Sage Publications. [Online] Available: doi: http://dx.doi.org.ez.sun.ac.za/10.4135/9781412950558. Accessed: 11 May 2015.

Mayne, J, 2001. Addressing attribution through contribution analysis: using performance measures sensibly. *The Canadian Journal of Evaluation*, **16**(1), 1-24.

Mouton, C. 2010. *The history of programme evaluation in South Africa*. Unpublished MPhil dissertation. Bellville: Stellenbosch University.

Mouton, J. 2007. Approaches to programme evaluation research. *Journal of Public Administration*, **42**(6), 490-511.

Naidoo, I.A. 2011. *The role of monitoring and evaluation in promoting good governance in South Africa: A case study of the Department of Social Development*. Unpublished PhD dissertation. Johannesburg: University of Witwatersrand.

Oliver, K., Lorenc, T. & Innvaer, S. 2014. New directions in evidence-based policy research: A critical analysis of the literature. *Health Research Policy and Systems*, **12**(34), 1-11.

Owen, J. & Rogers, P. 1999. *Programme Evaluation*. Sage Publications.

Parliamentary Monitoring Group (PMG). 2016a.*Question 2755/2013 for written reply*. *Ms E. More to ask the Minister of Social Development.* [Online] Available: https://pmg.org.za/question_reply/469/ Accessed: 10 October 2016.

Parliamentary Monitoring Group (PMG). 2016b. *Question 2322. Ms LL van der Merwe (IFP) to ask the Minister of Women, Children and People with Disabilities.* [Online] Available: https://pmg.org.za/question_reply/347/. Accessed 10 October 2016.

Parliamentary Monitoring Group (PMG). 2016c. *Question 1227. Mr M. Swart (DA) to ask the Minister of Finance. [*Online] Available: https://pmg.org.za/question_reply/86/. Accessed 10 October 2016.

Parliamentary Monitoring Group (PMG). 2016d. *Question 1725. The Leader of the Opposition (DA) to ask the Minister of Finance*. [Online] Available: https://pmg.org.za/question_reply/412/. Accessed 10 October 2016.

Parliamentary Monitoring Group (PMG). 2016e. *Internal question paper 1512012. Miss A.T. Lovemore (DA) to ask the Minister of Basic Education.* [Online] Available: https://pmg.org.za/question_reply/418/. Accessed 10 October 2016

Parliamentary Monitoring Group (PMG). 2016f. *Question 1056. Ms T. Gqada to ask the Minister of Human Settlements.* [Online] Available: https://pmg.org.za/committee-question/3416/. Accessed 10 October 2016.

Patton, M.Q. 1984. An alternative evaluation approach for the problem-solving training program**:** A utilization-focused evaluation process. *Evaluation and programme planning*, **7**(2), 189-192.

Patton, M.Q. 1986. *Utilization-focused Evaluation*. Thousand Oaks, CA: Sage Publications.

Patton, M.Q. 1990. *Qualitative Evaluation and Research Methods.* 2nd edition. Thousand Oaks, CA: Sage Publications.

Patton, M.Q. 1996. *Utilization-focused evaluation: The new century text.* 3rd edition. Thousand Oaks, CA: Sage Publications.

Patton, M.Q. 2008. *Utilization-focused Evaluation*. 4th edition. Sage Publications. Thousand Oaks, CA: Sage Publications.

Patton, M.Q. 2012. A utilization-focused approach to contribution analysis. *Evaluation,* **18**(3), 364-377.

Pawson, R. 2002a. Evidence-based Policy: In Search of a Method. *Evaluation,* **8**(2),157-181.

Pawson, R. 2002b. Evidence-based Policy: The Promise of Realist Synthesis. *Evaluation,* **8**(3), 340-358.

Pawson, R. 2002c. Evidence and Policy and Naming and Shaming. *Policy Studies,* **23**(3/4), 211-230.

Pawson, R. 2006. *Evidence-based policy: A realist perspective*. London: Sage Publications.

Pawson, R. 2013. *The science of evaluation: A realist manifesto*. London: Sage Publications.

Pawson, R. & Manzano-Santaella, A. 2012. A realist diagnostic workshop. *Evaluation*, **18**(2), 176-191.

Pawson, R. & Tilley, N. 1995. Whither (European) Evaluation Methodology? *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization,* **8**(3), 20-33.

Pawson, R. & Tilley, N. 1997a. *Realistic Evaluation*. London: Sage Publications.

Pawson, R. & Tilley, N. 1997b. An introduction to scientific Realist Evaluation. In Chelimsky, E. & Shadish, W.R. (eds.). *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: Sage Publications, 405-419.

Pawson, R. & Tilley, N. 2001. Realistic Evaluation Bloodlines. *American Journal of Evaluation*, **22**(3), 317-324.

Pawson, R. & Tilley, N. 2004. *Realistic Evaluation*. 1-36. [Online] Available: www.communitymatters.com.au/RE_chapter.pdf Accessed: 1 August 2016.

Pawson, R. & Tilley, N. 2005. Realistic Evaluation. In Mathison, S. (ed.). *Encyclopaedia of Evaluation*. Thousand Oaks, CA: Sage Publications, 363-368. [Online] Available: doi: http://dx.doi.org.ez.sun.ac.za/10.4135/9781412950558.n179. Accessed: 11 May 2015.

Pawson, R., Wong, G. and Owen, L. 2011. Known knowns, known unknowns, unknown unknowns: The predicament of evidence-based policy. *American Journal of Evaluation,* **32**(4), 518-546.

Pearce, W. & Raman, S. 2014. The new randomized controlled trials (RCT) movement in public policy: challenges of epistemic governance. *Policy Sci,* **47,** 387-402

Pedersen, L.H. & Rieper, O. 2008. Is Realist Evaluation a realistic approach for complex reforms? *Evaluation,* **14**(3), 271-293.

Pedersen, L.M., Nielsen, K. & Kines, P. 2012. Realistic evaluation as a new way to design and evaluate occupational safety interventions. *Safety Science,* **50**(1), 48-54.

Phillips, S., Goldman, I., Gasa. N., Akhalwaya, I. & Leon, B. 2014. A focus on M&E of results: An example from the Presidency, South Africa. *Journal of Development Effectiveness***, 6**(4), 392-406.

Picciotto, R. 1997. Evaluation in the World Bank: Antecedents, Instruments, and Concepts. In Chelimsky, E. & Shadish, W.R. (eds.), *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: Sage Publications, 201-215.

Porter, S. & Goldman, I. 2013. A Growing Demand for Monitoring and Evaluation in Africa. *African Evaluation Journal,* **1**(1), 10-18.

Porter, S. & O'Halloran, P. 2012. The use and limitation of realistic evaluation as a tool for evidence-based practice: A critical realist perspective. *Nursing Inquiry*, **19**(1), 18-28.

Punton, M., Vogel, I. & Lloyd, R. 2016. *Reflections from a Realist Evaluation in Progress: Scaling Ladders and Stitching Theory*. IDS.CDI Practise Paper #18, 1-11[Online] Available: http://www.ids.ac.uk/publication/reflections-from-a-realist-evaluation-in-progress-scaling-ladders-and-stitching-theory. Accessed 10 October 2016

Rabie, B. & Cloete, F. 2009. A new typology of monitoring and evaluation approaches. *Administratio Publica*, **17**(3), 76-97.

Rabie, B. 2011. *Improving the systematic evaluation of local economic development results in South African local government*. Unpublished PhD dissertation. Bellville: Stellenbosch University.

Republic of South Africa (RSA). 2007. *Policy framework for the Government-Wide Monitoring and Evaluation System (GWMES).* Pretoria: National Treasury.

Republic of South Africa (RSA). 2009. *Improving Government Performance: Our approach*. The Presidency: Department of Performance Monitoring and Evaluation (DPME).

Republic of South Africa (RSA). 2011a. *National Evaluation Policy Framework (NEPF).* The Presidency: Department of Performance Monitoring and Evaluation (DPME).

Republic of South Africa (RSA). 2011b. *Measuring success in human settlements development: An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa*. National Department of Human Settlements (NDHS). [Online] Available: http://microdata.worldbank.org/index.php/catalog/1041 Accessed: 22 June 2016.

Republic of South Africa (RSA). 2011c. *Child Support Grant Evaluation 2010: Qualitative Research Report.* Department of Social Development, SASSA and UNICEF. Pretoria: South Africa

Republic of South Africa (RSA). 2012. *The South African Child Support Grant Impact Assessment: Evidence from a survey of children, adolescents and their households*. Department of Social Development, SASSA and UNICEF. Pretoria: South Africa

Republic of South Africa (RSA). 2014. *Medium-Term Strategic Framework (MTSF).* The Presidency: Department of Planning, Monitoring and Evaluation (DPME).

Republic of South Africa (RSA). 2016a. *National Evaluation Plan 2016-17 to 2018-19.* 26 April. Department of Planning Monitoring and Evaluation (DPME). [Online] Available: http://www.samea.org.za/index.php?module=Pagesetter&type=file&func=get&tid=25&fid=file1&pid=2 Accessed: 18 February 2016.

Republic of South Africa (RSA). 2016b. *Media Statement: Release of a descriptive report detailing the progress of the Employment Tax Incentive.* The National Treasury: Pretoria.

Rogers, P. 2007. Theory-based evaluation: Reflections ten years on. *New Directions for Evaluation*, **114**(Summer), 63-67.

Rogers, P. & Peersman, G. 2014. Developing a research agenda for impact evaluation in development. *IDS Bulletin,* **5**(6). Oxford, United Kingdom: Institute of Development Studies.

Rogers, P., Petrosino, A., Huebner, T.A. & Hacsi, T.A. 2000. Program theory evaluation: Practice, promise, and problems. *New Directions for Evaluations*, **87**(Fall), 5-13.

Rossi, P. H. & Freeman, H. E. 1985. *Evaluation: A systematic approach*. 3rd edition. Beverly Hills, Sage Publications.

Rossi, P.H. & Freeman, H.E. 1993. *Evaluation: A systematic approach.* 5th edition. Newbury Park, CA: Sage Publications.

Rossi, P.H., Freeman, H.E & Lipsey, M.W. 1999. *Evaluation: A systematic approach.* 6th edition. Thousand Oaks, CA: Sage Publications.

Rycroft-Malone, J., Fontenla, M., Bick, D. & Seers, K. 2010. A realistic evaluation: The case of protocol-based care. *Implementation Science*, **5**(38).

Sayer, A. (2000). *Realism and social science*. London: Sage.

Scheyer, P. & Stake, R. 1976. Evaluation Reflections: A program's self-evaluation portfolio. *Studies in Educational Evaluation*, **2**(1), 37-40.

Scriven, M. 1997. Truth and Objectivity in Evaluation. In Chelimsky, E. & Shadish, W.R. (eds.), *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: Sage Publications, 477-501.

Scriven, M. 1999. The fine line between evaluation and explanation. *Research on Social Work Practice,* **9**(4), 521-524.

Scriven, M. 2003/2004. Michael Scriven on the differences between Evaluation and Social Science Research. *The Evaluation Exchange*, **IX** (4), Winter.

Segone, M. 2008. Evidence-based policy-making and the role of monitoring and evaluation within the new aid environment. In Segone, M., *Bridging the gap: The role of monitoring and evaluation in evidence-based policy-making*. Evaluation Working Paper #12. UNICEF, New York,16-45. [Online] Available: www.unicef.org/ceecis/evidence_based_policy_making.pdf Accessed: 1 August 2016.

Stake, R 1976. A theoretical statement of responsive evaluation. *Studies in Educational Evaluation*, **2**(1), 19-22.

Stake, R. 1991. Excerpts from: Programme evaluation, particularly responsive evaluation. *Evaluation Practice*, **12**(1), 63-76.

Stame, N. 2004. Theory-based evaluation and types of complexity. *Evaluation,* **10**(1), 58-76.

Statistics South Africa (StatsSA). 2010. *South African Statistical Quality Assessment Framework* (SASQAF). 2nd edition. Pretoria: Statistics South Africa.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. & Befani, B. 2012. *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*.UK Department for International Development. Working Paper 38. [Online]Availablehttps://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf Accessed: 1 November 2016.

Stufflebeam, D.L. 1966. A depth study of the evaluation requirement. *Theory into Practice*, **5**(3), 121-133.

Stufflebeam, D.L. 2001. Evaluation models. *New Direction for Evaluation,* **89**, 7-98.

Stufflebeam, D.L., Madaus, G.F. & Kellaghan, T. (eds.). 2000. *Evaluation Models.* Boston: Kluwer Academic Publishers.

Stufflebeam, D.L. & Shinkfield, A.J. 1985. *Systematic evaluation: A self-instructional guide to theory and practice*. Boston: Kluwer-Nijhoff Publishers.

Suchman, E.A. 1967. *Evaluative research: principles and practice in public service & social action programmes*. New York, NY: Russell Sage Foundation.

Talbot, C. & Talbot, C. 2015. Bridging the academic policy-making gap: practice and policy issues. *Public Money & Management*, **35**(3), 187-194. [Online] Available: doi: 10.1080/09540962.2015.1027491 Accessed: 1 August 2016.

Taylor, B. 2013. *Evidence-based policy and systemic change: Conflicting trends?* [Online] Available: http://ssrn.com/abstract=2304208 or http://dx.doi.org/10.2139/ssrn.2304208 Accessed: 1 August 2016.

United Kingdom. 2009. *Building the evidence to reduce poverty*. The UK's policy on evaluation for international development. Department for International Development.

Van der Berg, S., Girdwood, E., Shepherd, D., van Wyk., Kruger, J., Viljoen, J., Ezeobi, O. & Ntaka, P.2013. *The impact of the introduction of grade R on learning outcomes*. Research on socio-economic policy (ReSEP). Stellenbosch: Department of Economics, Stellenbosch University.

Weiss, C.H. 1990. Nuggets and Roots: The cliff notes version of my life in evaluation. In Alkin, M.C. *Debates on Evaluation*. Newbury Park, CA: Sage Publications.

Weiss, C.H. 1997a. Theory-based evaluation: past, present, and future. In Rog, D.J. & Fournier, D. (eds.). *Progress and Future Directions in Evaluation: Perspectives on Theory, Practice and Methods*. New directions for evaluation, No. 76. San Francisco, CA: Jossey-Bass, 41−55.

Weiss, C.H. 1997b. How can theory-based evaluation make greater headway? *Evaluation Review,* **21**(4), 501-524.

Weiss, C.H. 1997c. Theory-based evaluation: Past, present and future. *New Directions for Evaluation*, **76**(Winter), 68-81.

Weiss, C.H. 2000. Which links in which theories shall we evaluate? *New Directions for Evaluation*, **87**(Fall), 35-45.

Weiss, C.H. 2004a. Rooting for Evaluation: A cliff notes version of my work. In Alkin, M.C. (ed.). *Evaluation Roots: Tracing theorists' views and influences*. University of California, LA: Sage Publications, 154-169. [Online] Available: http://dx.doi.org.ez.sun.ac.za/10.4135/9781412984157.n9 Accessed: 11 May 2015.

Weiss, C.H. 2004b. On theory-based evaluation: Winning friends and influencing people. *The Evaluation Exchange*, **IX** (4). Harvard Family Research Project, Harvard Graduate School of Education.

Weiss, C.H., Murphy-Graham, E., Petrosino, A. & Gandhi, A. 2008. The fairy godmother and her warts: Making the dream of evidence-based policy come true. *American Journal of Evaluation*, **29**(1), 29-47.

Westhorp, G. 2014. *Realist impact evaluation: An introduction*. London: Methods Lab/Overseas Development Institute. [Online] Available: http://www.odi.org/publications/8716-methods-lab-realist-impact-evaluation-introduction Accessed: 2 February 2016.

Westhorp, G., Prins, E., Kusters, C., Hultink, M., Guijt, I. & Brouwers, J. 2011. *Realist Evaluation: An overview*. Seminar Report. [Online] Available: http://www.managingforimpact.org/sites/default/files/resource/2011_wp_realistevaluationseminar_cecilekusters_2x.pdf Accessed: 14 January 2016.

White, H. 2009. *Some reflections on current debates in Impact Evaluation*. International Initiative for Impact Evaluation (3ie), Working Paper. [Online] Available: http://www.3ieimpact.org/en/publications/working-papers/working-paper-1/ Accessed: 11 May 2015.

White, H. 2010. A contribution to current debates in Impact Evaluation. *Evaluation*, **16**(2), 153-164.

Wildschut, L.P. 2014. *Theory-based evaluation, logic modelling and the experience of SA non-governmental organisations*. Unpublished PhD dissertation. Bellville: Stellenbosch University.

Woolcock, M. 2013. *Using case studies to explore the external validity of 'complex' development* interventions. World Institute for Developing Economics Research (WIDER) Working Paper No 2013.096. United Nations University. [Online] Available: https://research.hks.harvard.edu/publications/getFile.aspx?Id=1007 Accessed: 1 August 2016.

Yin, R. 2009. How to do better case studies (with illustrations from 20 exemplary case studies). In Bickman, L. & Rog, D. (eds.). *The Sage Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage Publications.

APPENDIX A:

STATUS OF EVALUATIONS IN NATIONAL EVALUATION PLAN 2011-2015

Status of evaluations as at 30 September 2015

Name of Department	Title of evaluation	Status as at 30 September 2015
2011/12		
Social Development, Basic Education, Health	Diagnostic Review of Early Childhood Development	New policy gazetted
2012/13		
Trade and Industry	Implementation/ design evaluation of the Business Process Services Programme (BPS)	Final report approved by Cabinet. Improvement plan being implemented. Scheme relaunched
Basic Education	Impact Evaluation of Grade R	Final report approved by Cabinet. Improvement plan being implemented. Interventions to address quality.
Health (with Social Development, DAFF, DRDLR, DWCPD)	Implementation Evaluation of Nutrition Programmes addressing Children Under 5	Final report approved by Cabinet. Nutrition Plan being developed with Food Security.
Rural Development and Land Reform	Implementation Evaluation of the Land Reform Recapitalisation and Development Programme	Final report approved by Cabinet. Improvement plan being implemented.
Rural Development and Land Reform	Implementation Evaluation of the Comprehensive Rural Development Programme (CRDP)	Final report approved by Cabinet. Improvement plan being implemented.
Human Settlements	Implementation Evaluation of the Integrated Residential Development Programme (IRDP)	Evaluation underway.
Human Settlements	Implementation Evaluation of the Urban Settlements Development Grant (USDG)	Report approved. Changes made already to guidelines. About to be tabled at cabinet.
Basic Education	Impact Evaluation of the National School Nutrition Programme (NSNP)	Stopped and restarted in 2014/15.

Name of Department	Title of evaluation	Status as at 30 September 2015
2013-14		
Trade and Industry	Evaluation of Export Marketing Investment Assistance Incentive programme (EMIAI)	Final report approved by Cabinet. Improvement plan being implemented.
Trade and Industry	Evaluation of Support Programme for Industrial Innovation (SPII)	Final report approved by Cabinet. Improvement plan being implemented. Scheme relaunched.
Trade and Industry	Impact Evaluation of Technology and Human Resources for Industry Programme (THRIP)	Report approved. About to be tabled at cabinet.
Military Veterans	Evaluation of Military Veterans Economic Empowerment and Skills Transferability and Recognition Programme.	Report approved.
Science and Technology	Evaluation of National Advanced Manufacturing Technology Strategy (AMTS)	Stuck due to illness in evaluators. Stopped.
South African Revenue Services	Impact Evaluation on Tax Compliance Cost of small businesses	Draft report
Co-operative Governance	Impact evaluation of the Community Works Programme (CWP)	Report approved.
Rural Development and Land Reform	Evaluation of the Land Restitution Programme	Final report approved by Cabinet. Improvement plan being implemented.
Agriculture, Forestry and Fisheries	Impact Evaluation of the Comprehensive Agricultural Support Programme (CASP)	Report approved. About to be tabled at cluster and cabinet.
Agriculture, Forestry and Fisheries	Implementation Evaluation of MAFISA	Report approved. About to be tabled at cluster and cabinet.
Human Settlements	Setting a baseline for future impact evaluations for the informal settlements targeted for upgrading	Delayed by DHS procurement. Final report expected in October 2015.
Human Settlements	Evaluating interventions by the Department of Human Settlements to facilitate access to the city.	Delayed by DHS procurement and failure to get suitable SP. New appointment being made.
Human Settlements	Diagnostic of whether the provision of state-subsidised housing has addressed asset poverty for households and local municipalities	Delayed by DHS procurement. Draft report being revised.
Planning, Monitoring and Evaluation	Impact Evaluation of the Outcomes Approach	Major problems with implementation of the evaluation by the service provider. Stopped. Will restart at appropriate time for MTSF.

Name of Department	Title of evaluation	Status as at 30 September 2015
Presidency	Implementation Evaluation of Government's Coordination Systems	Final report approved by Cabinet. Improvement plan approved by FOSAD Manco June 2015.
Basic Education	Evaluation of the quality of the National Senior Certificate (NSC)	Dropped as a Ministerial Review underway
2014-15		
Environmental Affairs	Evaluation of the Effectiveness of Environmental Governance in the Mining Sector (EEGM)	Report approved. Being used as input for Mining Phakisa. Tabled soon at cluster/Cabinet.
Higher Education and Training	Design Evaluation of the Policy on Community Education and Training Colleges (PCETC)	Report approved. Changes made to policy. Tabled soon at cluster/Cabinet.
Human Settlements	Impact/Implementation Evaluation of the Social Housing Programme (SHP)	Draft report
Science and Technology	Evaluation of the Indigenous Knowledge Systems Policy (IKSP)	Underway
Social Development	Diagnostic Evaluation/ Programme Audit for Violence Against Women and Children (AVAWC)	Evaluation stopped due to inadequate performance. Will be readvertised.
Social Development	Diagnostic Review of the Social Sector Expanded Public Works Programme	Report approved, been to cluster and being tabled soon at Cabinet.
South African Police Service	Economic Evaluation of the Incremental Investment into the SAPS Forensic Services (SAPS)	Stuck – problems accessing information.
Agriculture, Forestry and Fisheries/ Rural Development and Land Reform	Implementation Evaluation of the Ilima Letsema Programme and cost-benefit analysis of the revitalisation of existing Irrigation Schemes	Dropped – due to delays carried over to 2015-16 and no budget.
Agriculture, Forestry and Fisheries	Impact evaluation of MAFISA (quantitative) including establishing a baseline	Evaluation stopped following termination of the programme by Treasury. Instead doing impact evaluation of Extension Recovery Programme
Agriculture, Forestry and Fisheries, with the Department of Rural Development and Land Reform	Policy Evaluation of Small Farmer Support	Underway.
Basic Education	Evaluation of the Funza-Lushaka Bursary Scheme	Report approved. Awaiting management response and improvement plan.

Name of Department	Title of evaluation	Status as at 30 September 2015
Basic Education	Implementation Evaluation of the National School Nutrition Programme	Draft report received.
Rural Development and Land Reform	Impact evaluation of Land Restitution Programme (quantitative) including establishing a baseline	Service provider selected. 3ie managing evaluation. Treasury secured additional funding to enable a 7 year impact study.
Planning, Monitoring and Evaluation	Impact/implementation evaluation of the MPAT system	Cabinet has approved report. Improvement plan being implemented
Planning, Monitoring and Evaluation	Impact/implementation evaluation of the Strategic Planning/APP system	Service provider appointed.
2015-16		
Agriculture, Forestry and Fisheries	Agricultural Extension Recovery Plan	Underway
Basic Education	Evaluation of CAPS/New School Curriculum	Underway
National Prosecuting Authority	Evaluation of the Asset Forfeiture Unit Sub-programme	Stuck due to changes in NPA.
Social Development	Diagnostic evaluation of the Non-Profit Organisations Regulatory Framework and Legislation	Underway
Social Development	Implementation Evaluation of the National Drug Master Plan in addressing all forms of Substance abuse	Service provide appointed
Higher Education and Training	Evaluation of the National Qualifications Framework Act (NQFA)	TORs not yet finalised
Basic Education	Evaluation of Early Grade Reading in SA	Underway
Mineral Resources	Implementation evaluation of the mining charter	Dropped as having Operation Phakisa on mining
Public Service and Administration	Service Delivery Improvement Planning System	TORs not finalised
Planning, Monitoring and Evaluation	Implementation evaluation of citizen-based monitoring (CBM)	Draft report received.
Planning, Monitoring and Evaluation	Impact/implementation evaluation of the evaluation system	Delayed to 2016/17 as insufficient budget

Source: RSA, 2016a: 15-17.

APPENDIX B:

SUMMARY OF APPROVED EVALUATIONS IN NATIONAL EVALUATION PLAN 2016/17

Name of Department	Intervention to be evaluated	Key motivation for this evaluation including scale (eg budget, beneficiaries)
Higher Education and Training	Evaluation of the Technical and Vocational Education and Training (TVET) Colleges Expansion and Capacity Development Programme	The National Skills Fund (NSF) was established in 1999 in terms of the Skills Development Act, 1998 (as amended). In October 2011, the Minister of Higher Education and Training requested the NSF to release a portion of the surplus (uncommitted) funds from the Discretionary Projects Programme to support the expansion of student enrolment in TVET Colleges. Following the Minister's request, the NSF set aside a budget of R2.5 billion for the College Expansion Programme. It is expected that approximately 102,000 more students/learners will access learning opportunities at TVET Colleges through the above-mentioned NSF allocation of funding over the 3-year period (2012-2015). The College Expansion Programme has four main objectives; to support TVET Colleges to enrol a larger number of students, particularly in artisanal-related programmes and programmes that would support the needs of the economy; to support TVET Colleges to offer a broader set of programmes through an expanded programme-qualification mix; to support TVET Colleges to expand student access to workplace-based learning (WBL) through learnerships (including artisanships) and internships; and d) to support TVET Colleges to build the capacity of their staff. The evaluation will therefore inform the management structures of the DHET and the NSF whether to continue funding similar programmes such as the TVET College Expansion and Capacity Development Programme. This evaluation is important because it will inform the management structures of the DHET and the NSF whether to continue funding similar programmes such as the TVET College Expansion and Capacity Development Programme.
Justice	Implementation/ Design Evaluation of the Integrated Justice System / Programme	The IJS derives its mandate from the Justice Crime Prevention and Security (JCPS) Cluster's Strategy, which in turn derives its mandate from Outcome 3 (All People in South Africa are and feel safe) of the Delivery Agreement with the South African Government. The Medium-Term Strategic Framework (MTSF) Sub-outcome 2 {An Efficient and Effective Criminal Justice System(CJS)} seeks to realize Vision 2030 of the National Development Plan (NDP) Chapter 12 – Building Safe Communities, that requires that people living in South Africa feel safe at home, at school and at work and that they enjoy a community life free of fear. The implementation of the IJS programme is intended to provide South Africa with a world-class integrated criminal justice system that will address system blockages such, non-existence of functional and business integration amongst JCPS departments, policy misalignment, lack of timely access to criminal record history and notification of events, imbalances in the level of automation of departments and incompatible information technology platforms, and a lack of quality information and information sharing. Estimated total budget for the intervention (over 3 year MTEF period) is R1,5 billion.

Name of Department	Intervention to be evaluated	Key motivation for this evaluation including scale (eg budget, beneficiaries)
Social Development	Implementation Evaluation of Older Persons Act	The Department of Social Development has developed the Older Persons Act (Act 13 0f 2006) to protect and empower older persons. The Act and its Regulations came into effect on the 1st of April, 2010 and calls for a developmental approach in dealing with ageing issues. The objective of the Act is to deal effectively with the plight of older persons by establishing a framework aimed at the empowerment and protection of older persons and at the promotion and maintenance of their status, rights, well-being, safety and security; and to provide for matters connected therewith. The aim of the Older Persons Act is to move services from institutional care to community-based care and support services. There are 4.5 million Older Persons in South Africa and this is about 8.3% of the population. The 2014/15 financial year budget for the intervention was R13 million for DSD national department and R1.1 billion for the DSD Provincial department. This evaluation is important and timely because the DSD is in the process of amending the Act and the findings of the evaluation will therefore strengthen the implementation of the amended Act.
National Treasury	Evaluation of City Support Programme (CSP)	The CSP is a demand-driven and “umbrella” programme that covers a range of support mechanisms for metropolitan municipalities and the broader intergovernmental environment to contribute to the achievement of well governed, inclusive, productive, sustainable cities. These mechanisms are implemented in three main areas • Creating an enabling intergovernmental environment for city transformation through changes in the policy and regulatory environment • Restructuring the fiscal and financial framework for the cities • City implementation support. Providing and integrated package of implementation support to cities The CSP responds to the challenges of cities in South Africa - they are insufficiently productive, segregated and exclusionary, unsustainable and governance is fragmented and failing to deal with the legacy of apartheid. Productive, inclusive and sustainable cities are pre-requisites for urban economic growth and a reduction in poverty and inequality. The budget for intervention (not for the evaluation) for 2014/15 financial year is R225 million (R75 million in technical support and R150 million on grant support to cities for expenditure in integration zones), ie approximately R700 million over the MTEF period.
Home Affairs	Evaluation of Birth Registration Programme	The early registration of birth is essential to ensure the integrity of the National Population Register, which is used to affirm the identity and status of citizens and gives them access to rights and services. The focus of the birth registration programme is to ensure that all births in South Africa are registered within 30 days to secure identity, civil status and the National Population Register. Under colonialism and apartheid the births of Africans in particular were not systematically captured, necessitating a Late Registration of Birth (LRB) procedure that led to widespread fraudulent acquisition of identities and citizenship. The number of children registered within the legislated 30 days of birth has increased from 46% in 2010-2011 to 64% in 2014-2015. A significant development is the replacement of the abridged birth certificate with a full birth certificate that can be printed in frontline offices. The capturing of both parents details to secure the identity of the child is an important feature and is now a requirement for any minor travelling through a Port of Entry. Amendments to Legislation have also drastically increased penalties for identity and vital registration fraud. Another critical development is the ending of the current LRB regime in December 2015 and replacing it with much more rigorous requirements and escalating penalties for those who register births after 30 days. Of particular concern are the poor, the vulnerable, and the marginalised, rural and institutionalised populations. The evaluation will assess how the system is working and how it can be strengthened.

Name of Department	Intervention to be evaluated	Key motivation for this evaluation including scale (eg budget, beneficiaries)
Science and Technology	Design and Implementation Evaluation of the National Space Strategy	South Africa needs to become a competitor globally and an exporter of technology in order to increase country's share in the global space and be an independent country. Space applications have already been useful from an overall societal perspective, and could be of further assistance in addressing major societal challenges over the coming decade. The five major societal challenges that space can address are related to the state of the environment, the management and use of natural resources, the increasing mobility of individuals and products and its consequences, growing security threats, and the shift towards the information society. The Department of Science and Technology resuscitated the national space programme in 2005. The policy and strategy of the programme were approved by 2008. The findings of the evaluation will enhance the DST's understanding on the implementation and effectiveness of the strategy through assessing the: • Support on targeted training and awareness programmes which promotes skills development in key space science and technology areas (satellite engineering and space applications development) for academia, public and private industry and also contributes towards outcome one and five. • Contribution of the Department in supporting manufacturing of satellite by upgrading existing and the development of new space infrastructure which contributes towards outcome six, ten, and eleven. • Contribution of the DST on acquiring, processing and distribution of space – based data for the development of space applications (products and services) for use by national and provincial government departments, parastatals, science councils, Non – government organisations and tertiary education institution which contributes to outcome two, three, seven, eight, 10 and 11. The budget for the intervention for the 2014/15 financial year is R 355 million with an estimated total budget for the intervention (over 5 year MTEF period) of R 8 billion.
Planning, Monitoring and Evaluation	Evaluation of the evaluation system	The Department of Performance Monitoring and Evaluation in the Presidency was established in April 2010. The initial rationale for the Department was the establishment of 12 priority outcomes, development and monitoring of plans against those priority outcomes. In 2011 DPME also started to develop the concept for a National Evaluation System, and a National Evaluation Policy Framework was approved by Cabinet on 23 November 2011. The basic evaluation system is now fairly well established based on National Evaluation Plans with 47 evaluations completed or underway covering around R75 billion of government expenditure (MTEF). The evaluation will look at the uptake of evaluation results, how the systems are working and proposed ways of strengthening it.
Environmental Affairs	The Impact of the Environmental Impact Assessment (EIA) Regime of sustainable development	NEMA's environmental impact management regime, particularly the process known as the Environmental Impact Assessment (EIA) process, has been South Africa's key regulatory instruments to mitigate and/or manage the impacts of new developments and activities that are considered to potentially undermine everyone's right to an environment that is not harmful to health and well-being. The evaluation is important as it will assess the effectiveness and credibility of the EIA process and its contribution toward sustainable development. The Department of Environmental Affairs is in the process of reviewing the EIA Regulations and developing a Coordinated Integrated Permitting System. This evaluation will inform the refinement of the Regulations and system development. The 2013/14 financial year budget for the intervention amounts to hundreds of million rands for the State and the developers. Around 100 000 South Africans participate directly in EIAs every year.

Name of Department	Intervention to be evaluated	Key motivation for this evaluation including scale (eg budget, beneficiaries)
Dti, Treasury, plus others	Evaluation of Business Incentives	A number of evaluations have been conducted of specific incentives schemes – ranging from dti incentives such as Support Programme for Industrial Innovation (SPII), SARS (Tax reform for small businesses), DRDLR - Land Recapitalisation and Development, to microfinance by DAFF (MAFISA). However, these have been conducted as separate evaluations without a bigger picture of the incentives environment and how these add together. This year more than R7 billion will be transferred directly from the fiscus to private companies to support business activity. In addition, the government foregoes at least R25 billion in revenue each year as part of various tax incentives to the private sector. The Ministers' Committee on the Budget has requested a detailed assessment of the impact of business incentives on economic growth, productivity, empowerment, competitiveness, the balance of trade and employment creation.

Source: RSA, 2016a: 11-13.

APPENDIX C:

SUMMARY OF PROPOSED EVALUATIONS IN NATIONAL EVALUATION PLAN 2017/18

Table 4: Summary of proposed evaluations for 2017/18

Name of Department	Name of intervention	Title of evaluation	Key motivation for this evaluation including scale (eg budget, beneficiaries)
Basic Education (DBE)	Kha Ri Gude Programme	Evaluation of Kha Ri Gude Programme	The Kha Ri Gude Mass Literacy Campaign was launched in February 2008, with the intention of enabling 4,7 million adults above the age of 15 years to become literate and numerate in one of the eleven official languages. Achieving this goal will enable South Africa to reach its UN: Education For All commitment made at Dakar in 2000 - that of halving the country's illiteracy rates by 2015. The Campaign makes specific efforts to target vulnerable groups. Currently 80% of the learners are women, 8% are disabled and 25% are youth, and 20% are above the age of 60. The total national and provincial budget for 2014/15 was a total of R484 million.
Department of Public Enterprises	Competitive Supplier Development Programme	Implementation evaluation of the Competitive Supplier Development Programme	In 2007, the CSDP was launched, however at the time, the procurement capability of State Owned Companies was not geared for supplier development. SOC were only focused on short term, transactional procurements which were often crisis driven. There was institutional fragmentation which meant that demand was not aggregated and the focus on a balance sheet perspective further limited the SOC planning horizon. In 2011, the supplier development was integrated into the SOC procurement policy, in particular for Eskom and Transnet. Subsequently, Transnet initiated planning for locomotive fleet procurement while Eskom initiated planning for filter bag procurement based on massive industrial impact. Between 2012 and 2013, the Eskom and Transnet's next generation supplier development plans were assessed and approved. Thereafter, the Supplier Development Plans (SDP) for both Eskom and Transnet were launched. Considerable progress has been made to position the SOC to support customers in emerging industrial sectors. Diagnostic studies were undertaken in both the automotive sectors and for port-related industries.

Source: RSA, 2016a: 31.

APPENDIX D:
SAMPLE ACCESS LETTER

PO Box 610
Bellville 7535
South Africa
Tel: +27 21 9184400
Fax: +27 21 9184468
E-mail: djam@sun.ac.za
Web: www.spl.co.za

19 September 2016

To Whom It May Concern

In pursuing the national objective of promoting research that would benefit practice and society, the School of Public Leadership of Stellenbosch University, requires of its senior postgraduate students to apply the theoretical knowledge gained in classes to a real-world issue that may be solved or at least result in new insights gained from the research they have to do.

We therefore kindly request that you grant our student, **Ms. P.N. Mbava (14986477)** an interview and/or access to information in your department with a view of answering a research question related to adopting a realist evaluation approach in policy impact evaluation. All information that is provided by you will be governed by the ethical research practices of the University. If deemed necessary, the research result, whether it is a research report or thesis, may be classified as confidential in order to restrict public access to such documents.

We believe that the research will be of value to the evaluation community and trust that it will be of relevance to you as well. We would be pleased to send you a copy of the research document once finalised you should you so require. Should you require any further information, please contact me directly.

Yours sincerely,
Babette Rabie
PhD Supervisor

DR BABETTE RABIE
Senior Lecturer in Public Policy and Public Sector Evaluation/
Senior Lektor in Openbare Monitering, Evaluering en Beleid
BA, BA (Hons), MA, PhD

School of Public Leadership
Stellenbosch University
Bellville Park Campus
Carl Cronje Drive
PO Box 610 Bellville
7535 South Africa

Skool vir Publieke Leierskap
Universiteit Stellenbosch
Bellvilleparkkampus
Carl Cronje Rylaan
Posbus 610 Bellville
7535 Suid-Afrika

T +27 (0)21 918 4186 / 918 4122 (Reception/Ontvangs)
F +27 (0)21 918 4123
babette.rabie@spl.sun.ac.za www.spl.sun.ac.za

SPL
SCHOOL OF PUBLIC LEADERSHIP
SKOOL VIR PUBLIEKE LEIERSKAP

APPENDIX E:

QUESTIONNAIRE TO POLICY-MAKERS

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

IMPORTANT NOTICE

You are kindly requested to complete this questionnaire for the purpupose of assisting the researcher, to fufill the requirements of the Phd in Public and Development Management at Stellenbsoch University.

The results of the study will inform the evaluation of strategic government programmes in the basic education, social protection and social housing sectors and will be of value to the evaluation community.

It should take approximately 20 minutes to complete this questionnaire

As a participant in this study your right to privacy, anonymity and confidentiality will be honoured at all times. Participants will not be identifiable in any way from the result of the study.

The names of participants will not appear in the report without the written consent of each individual.

A summary of the findings and final report of the research will be provided to all participants.

A consent forms for participating in this study will be provided to participants to complete.

Your time and attention in contributing to this study is sincerely appreciated.

Please complete the questionnaire by **Thursday 29 September 2016**

Respondent Name: _____

Respondent Title: _____

Respondent Business Address: _____

Rig asseblief alle korrespondensie aan die Registrateur/Please address all correspondence to the Registrar
Universiteitskantoor/University Office
Privaatsak/Private Bag X1 • Matieland, 7602 • Suid-Afrika/South Africa, Faks/Fax: +27 (0) 21 808 3800

1. As a policy-maker, which of the following aspects do you 'most expect /need' from a policy evaluation? Please tick all that is relevant to you:

Kindly cross (X) an appropriate box to indicate your choice.

1.1		What were the outcomes / results of the policy evaluation
1.2		The problems encountered in implementing the policy
1.3		How the policy was implemented
1.4		Any unexpected programme outcomes
1.5		Who benefited primarily from the policy
1.6		If the intervention was successful, when and where can it be replicated
1.7		The impact of demographic effects like culture on the expected programme outcomes
1.8		Who did not benefit from the policy
1.9		Why some do and others do not benefit from the policy
1.10		The theory assumptions on how change is supposed to happen
1.11		The budget
1.12		Other (please describe) –

2. Referring to the list above, in your opinion, what would policy-makers deem to be the 5 most meaningful, valid and useful important aspects in this list. **Please rank this in the order of importance, with 1 being the most important aspect.**

2.1

_____

2.2

_____

2.3

_____

2.4

_____

2.5

_____

3. Why do you think policy-makers would regard these aspects to be most important? Please provide a brief motivation for each ranked aspect.

3.1

_____

3.2

_____

3.3

_____

3.4

_____

3.5

_____

4. I would imagine that these expectations may vary between policy-makers and evaluations. Do you think all policy-makers would agree with your ranking above?

Kindly cross (X) an appropriate box to indicate your choice

4.1	Yes	
4.2	No	

4.2.2 If not, please explain possible factors that may explain these differences?

5. In your experience, what are the most important gaps/limitations with existing policy impact evaluations?

6. To what extent are you familiar with *The impact of the introduction of Grade R on learning outcomes/The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households / Youth wage subsidy experiment for South Africa/ An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa Evaluations?*

7. Which aspects of *The impact of the introduction of Grade R on learning outcomes/The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households / Youth wage subsidy experiment for South Africa/ An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa Evaluation* were you involved with?

8. What were the most valuable contributions of the findings of this evaluation to policy decision-makers?

9. In your opinion, what were the most important gaps/limitations of this evaluation to policy decision-makers?

10. Following the evaluation what were some of the questions raised by other policy-makers that you are aware of.

Interviewer offers a brief explanation of Realist Evaluation approach and the potential value of the approach on impact evaluations.

Realist Evaluation is a theory-driven evaluation approach. The approach seeks to understand the context under which a programme is implemented because it is believed such a context has an influence on how the programme achieves change. Both the context and how the programme achieve change results in the observed programme outcomes. Realist Evaluation has an explanatory focus and go beyond asking, 'What works?' or 'Does this programme work?' but rather ask, 'What works for whom in what circumstances and in what respects, and how?' This is precisely because "programmes never work indefinitely, in the same way, or in all circumstances, nor do they work for all people" (Pawson & Tilley, 1997).

11. In terms of your earlier responses, do you think adopting such an approach would offer useful, meaningful or valid results to policy decision-makers, perhaps beyond what they derived from the existing *The impact of the introduction of Grade R on learning outcomes/The South African child support grant impact assessment: Evidence from a survey of children, adolescents and their households / Youth wage subsidy experiment for South Africa/ An impact evaluation study of the Upgrading of Informal Settlements Programme in selected projects in South Africa evaluation?*

Kindly cross (X) an appropriate box to indicate your choice

11.1	Yes	
11.2	No	

12. Please explain why?

13. What would you regard as potential drawbacks or negative implications of adopting a Realist Evaluation approach?

14. Do you think the evaluation approaches adopted in evaluations in general in the public sector are always appropriate to inform the needs of policy decision-makers?

15. Are there any other factors that you think I should know, to really understand what makes policy evaluations useful, meaningful and valuable to policy-makers?

Thank you for your participation

APPENDIX F:
QUESTIONS ASKED BY POLICY-MAKERS: THE SOUTH AFRICAN CHILD SUPPORT GRANT

QUESTION 2755/2013

FOR WRITTEN REPLY

Date of publication on internal question paper: 18 October 2013

Internal question paper no: 33

2755. Ms E More (DA) to ask the Minister of Social Development:

(1) During the re-registration for grants drive, how many people were found to be receiving a child support grant, despite the fact that they were not the primary caregiver of the child and the child was in fact living elsewhere;

(2) what did her department do in these cases to ensure that the child support grant was transferred to the actual caregiver and not simply suspended;

(3) how many child support grants were suspended during the re-registration drive? NW3255E

REPLY:

(1) For the period ending September 2013 a total of 181,470 CSG grants were lapsed owing to a variety of reasons, amongst these, being those found not to be the primary care givers of the children in respect of whom they were collecting a child support grant.

(2) New grant applications were processed in the name of the new primary care-giver in whose physical care these children were found.

(3) At the end of September 2013 a total of 181,470 child support grants were lapsed.

Reply received: November 2013

Source: PMG, 2016a.

QUESTION 2322

Ms LL van der Merwe (IFP) to ask the Minister of Women, Children and People with Disabilities:

(1) What steps does she intend to take with regard to the United Nations Children's Fund (UNICEF) report that more than half of children in South Africa live in poverty and that one in four is HIV-positive;

(2) Whether she will make a statement on the matter?

REPLY:

(1) The UNICEF report shows a substantial drop in child poverty since 2003. This is further confirmed by analysis from the Children's Institute using StatsSA's General Household Survey data from 2003 to 2012. Using a poverty line equivalent to 2 dollars a day, child poverty dropped by about 19 percent points between 2003 and 2012. Progressive policies, in particular the expansion of the country's social assistance programmes largely accounts for this drop in child poverty. It is also noteworthy that many vulnerable families and their children were assisted to avoid substantial decline in their living standards during the global economic crisis of 2008 – 2009. A study by the Financial and Fiscal Commission and UNICEF, and in partnership with Stellenbosch University, confirmed that without the Child Support Grant, child poverty would have increased by some 9 percentage points during the economic recession. On HIV, the UNICEF report presents Department of Health and UNAIDS data to show the substantial progress in preventing mother – to – child transmission of HIV since 2004. The number of HIV positive pregnant women receiving anti-retrovirals increased from an estimated 32, 5000 in 2004 to 250 100 in 2010. This resulted in a substantial drop in new child infections during the period. Furthermore, a recent impact evaluation by MRC with support from CDC and UNICEF show that progress that has been made over the past decade in the implementation of the national PMTCT programme has enabled South Africa to reduce mother – to – child transmission of HIV to an estimated 2.7% at six weeks after birth-including 1.98% in the Western cape to 3.8% in Eastern Cape and Free State;

(2) No.

Reply received: September 2012

Source: PMG, 2016b.

APPENDIX G:
QUESTIONS ASKED BY POLICY-MAKERS: YOUTH WAGE SUBSIDY EXPERIMENT FOR SOUTH AFRICA

QUESTION NUMBER: 1227

DATE OF PUBLICATION: 11 SEPTEMBER 2009

Mr M Swart (DA) to ask the Minister of Finance:

(1) (a) When was the pilot programme on the youth wage subsidy initiated, (b) how much money has already been spent on the programme in each month by (i) actual wage subsidy payout, (ii) monitoring costs and (iii) other costs, (c) how many persons are participating in the programme and (d) when will the scheme be implemented; (2) whether the National Treasury has any estimates on the (a) cost of implementing the youth wage subsidy and (b) number of persons to participate within the first year of its implementation; if not, what is the position in this regard; if so, what are the relevant details?

REPLY:

(1) The youth wage subsidy is not a pilot programme, but a research project undertaken by a university-based research team into the effect of a youth wage subsidy on employment. The National Treasury in conjunction with the Department of Labour, is funding this research project.

(1) (a) The first phase of this research project began in April 2009 with the process of building the sample of youths who are to be part of the research project. This phase was completed at the end of August.

(1) (b) The allocation and distribution of subsidies among the sample has yet to take place. As such no money has been spent on (i) actual wage subsidy payout or (ii) monitoring costs for the research experiment. (iii) As of August 13, 2009, an amount of R465 515 had been spent and a further R143 077 committed towards the individual surveys of the sample, overhead costs and salaries of the university-based research team.

(1) (c) There are 4 000 youths aged between 20 and 24 participating in the research experiment.

(1) (d) There is no scheme to be implemented as this project is only a research experiment at present.

(2) No, it is not possible to provide estimates for (a) the costs of implementing a youth wage subsidy or (b) the number of persons to participate in the first year, as there is no policy decision to introduce a youth wage subsidy in South Africa.

Source: PMG, 2016c.

QUESTION NUMBER 1725 [NW2108E]

DATE OF PUBLICATION: 27 JULY 2012

The Leader of the Opposition (DA) to ask the Minister of Finance:

(1) Whether he intends transferring funds earmarked for the implementation of the Youth Wage Subsidy in the national Budget in 2011 to provinces that are willing to support the implementation of the policy; if not, why not; if so, what are the relevant details;

(2) whether he intends to provide any additional assistance to provincial governments that have implemented policies that aim to subsidies the employment of young South Africans; if not, why not; if so, what are the relevant details?

NW2108E

REPLY:

(1) The proposed youth employment incentive would be run through the South African Revenue Service (SARS) as a tax incentive to firms that hire young, inexperienced workers. As a tax expenditure, there is no allocated budget for the youth employment incentive. The estimated cost to the fiscus of approximately R5 billion over three years will be through foregone tax revenue, not earmarked funds, and the actual amount will depend on the uptake and job creation for young people that takes place due to the incentive.

Leveraging an existing operational and administrative platform confers significant benefits over developing piece-meal and new operational platforms to administer

Reply received: September 2012. Reply received: July 2012

Source: PMG, 2016d.

APPENDIX H:

NATIONAL TREASURY STATEMENT ON YOUTH WAGE SUBSIDY

national treasury
Department:
National Treasury
REPUBLIC OF SOUTH AFRICA

MEDIA STATEMENT

Release of a descriptive report detailing the progress of the Employment Tax Incentive

The Employment Tax Incentive (ETI) was introduced by Government on 1 January 2014 as part of a package of programs to address the social and economic problem of youth unemployment. The incentive aims to stimulate employment of 18 to 29 year olds in the formal sector by reducing the risks and costs associated with hiring younger workers, who tend to be inexperienced.

Youth employment is a critical component of the overall unemployment challenge. The burden of this unemployment falls heaviest on the poorest in society, with severe consequences for both the youth and the economy.

The ETI was legislated to continue until 31 December 2016, and if no legislative amendments are made before the end of the year then the ETI will expire. Parliament had requested a full review of the incentive in order to inform whether it should be continued, refined or allowed to lapse.

Government is currently engaging on these issues with other constituencies in NEDLAC. The inputs for the review to assess the impact of the incentive will also include feedback from constituencies and affected parties and additional independent research using the tax data. If amendments are required to extend or alter the design of the ETI, a draft version of the legislation will be published in the next two months to allow for public comment and sufficient time for the legislation to be considered in Parliament. The current draft tax bills before Parliament therefore do not deal with the ETI since the review is still in progress.

The report makes use of a unique dataset made available to National Treasury by the South African Revenue Service for the purposes of policy evaluation. The data is only available for the 2013/14 and 2014/15 tax years, due to the lags in tax data reporting. The dataset has not been used in this manner for detailed policy evaluation before, and as a result, significant attention is devoted to explaining the dataset in order to inform policy-makers, social partners and interested members of the public.

The data suggest that take up of the ETI has been strong. R6.06 billion was claimed between 1 January 2014 and 31 March 2016. In 2014/15, 32 368 firms lodged at least one claim on the ETI. Whilst this is a large number of firms, this represents 15% of firms in the tax database with eligible employees. The ETI has been claimed for 134 923 jobs in 2014 and 686 402 jobs in 2015. This implies the ETI supported approximately 5 per cent of all jobs in the tax dataset based on individual employee tax certificates in the 2014/15 tax year.

In the 2011 discussion document, it was estimated that R5bn would be spent on the youth wage subsidy over three years, supporting 423 000 jobs, of which 178 000 would be new jobs or jobs saved from loss. Since 2011, changes to the timing of spending, as well as slight differences in design have occurred, but broadly the estimates of jobs supported are higher than the initial 2011 projections.

As with all incentive evaluations, changes in the external environment make it harder to assess whether the ETI created new jobs or prevented a further worsening in youth unemployment. It is not possible to use descriptive data to determine whether these supported jobs are new jobs created, jobs that have been saved from being lost or jobs that would have been created anyway. To make this estimate, it is necessary to make assumptions about what might have happened in the absence of the ETI. Econometric studies provide a tool for doing this, but are beyond the scope of this paper.

These reports will be made available to the public as they are completed.

Issued by National Treasury on 26 August 2016

Source: RSA, 2016b.

APPENDIX I:
QUESTIONS ASKED BY POLICY-MAKERS: THE IMPACT OF INTRODUCTION OF GRADE R ON LEARNING OUTCOMES

DATE OF PUBLICATION FOR INTERNAL OUESTION PAPER: 02/06/2012

INTERNAL QUESTION PAPER 1512012

Miss AT Lovemore (DA) to ask the Minister of Basic Education:

(1) (**a) Why have Grade R teachers not been incorporated in the staff establishment of provincial departments of education?**

The Education White Paper 5 on Early Childhood Development (2001) proposed a transition of providing grade R in the system through grants in aid by provincial departments of education to School Governing Bodies. This makes ECD teachers the employees of the governing bodies.

(b) **Why have Grade R teachers not been recognised as teachers by the SA Council for Educators (SACE)?** and

The minimum qualification required for reaching in a grade R class is an ECD level 4 on the National Qualification Framework. The department has reached an agreement with the SA Council for Educators to allow for the conditional

(c) When is it anticipated that Grade R teachers will be incorporated in the provincial departments of education and recognised by SACE?

The minimum qualification for employment as an educator in the system is a grade 12 plus 4 years degree. The National Norms and Standards for Grade R Funding of 2008 (section 182 (a) allows provinces to establish posts to support the funding of grade R in public schools. Section 182jb) allows provinces to convert a portion of a schools total allocation to grade R to a positions taking into account the total cost of the post to the state. Provinces are aware of this policy position.

(2) **Whether standardised training is available for aspirant Grade R teachers; if not, why not; if so, what are the relevant details?**

There are public and private FET colleges accredited to offer the ECD level 4 and 5 qualifications. Higher Education Institutions also offer the ECD level 5 towards a Bachelor

of Education in foundation Phase. There is a new grade R certificate that has been introduced .as a bridge to allow ECD practitioners access to a degree for employment purposes with effect from 2013.

(3) **What institutions offer training to Grade R teacher in each of the provinces**

(See Annexure A for details per province)

(4) **Whether all public schools offer enrolment in Grade R**

a) No. There are 22,782 public schools offering grade R.

b) Why not

Grade R learners are also enrolled at community ECD centres. This is catered for in the provisioning model in the ECD White Paper (SOT6 public and 20% community)

(b) **What is the extent of Grade R coverage for schools in each of the provinces**? NW1762E

Grade R coverage as at end of 2011 is as follows:

1. Eastern Cape - 157 184
2. Free State - 28 627
3. Gauteng - 86 240
4. KwaZulu-Natal - 181 585
5. Limpopo - 117 279
6. Mpumalanga - 56 726
7. Northern Cape - 13 153
8. North West - 42 937
9. Western Cape - 50923

Total - 734 654

(Taken from the School Realities 2012)

Here is Annexure A: www.pmg.org.za/files/questions/RNW1484-120704.pdf

Reply received: July 2012.

Source: PMG, 2016e.

APPENDIX J:
QUESTIONS ASKED BY POLICY-MAKERS: UPGRADING OF INFORMAL SETTLEMENTS PROGRAMME

MINISTRY FOR HUMAN SETTLEMENTS

REPUBLIC OF SOUTH AFRICA

NATIONAL ASSEMBLY

QUESTION FOR WRITTEN REPLY

QUESTION NO.: 1056

DATE OF PUBLICATION: 11 APRIL 2016

Ms T. Gqada (DA) to ask the Minister of Human Settlements:

(1) Whether her department uses a standard checklist for the upgrading of informal settlements; if not, why not; if so,

(2) whether she will provide Ms T Gqada with a copy of the specified checklist;

(3) (a) how many informal settlements have been upgraded (i) in the (aa) 2012-13, (bb) 2013-14, (cc) 2014-15 and (dd) 2015-16 financial years and (ii) since 1 April 2016 as part of her department's informal settlements upgrade programmes, (b) what are the names of these settlements, (c) where is each specified settlement situated and (d) which services did each specified settlement receive during its upgrade? NW1189E

REPLY:

(1) The National Housing Code contains a comprehensive set of guidelines for the upgrading and development of informal settlements. The guidelines assist human settlements and housing practitioners in planning, funding and implementation of the approved informal settlements upgrading policy and programme. It is to be noted that the guidelines contained in the National Housing Code have been bench-marked and aligned to international good practice and that component parts of the South Africa policy, funding and implementation good practice, have been incorporated into international and national country policies and programmes.

In terms of the Upgrading of Informal Settlements Programme (UISP), informal settlement upgrading should be undertaken in phases, with Phases 1 to 3 focusing on community participation, supply of basic services and security for all residents. The priority is to address issues of household health and safety including the provision of interim services as a minimum norm and standard, in the form of reasonable access to water, sanitation, storm water management and road access to households. The current upgrading approach is incremental and infrastructure-led, and recognises that meaningful developmental improvements need to be provided to all informal settlements as rapidly as possible. Upon a settlement having been formalised in the form of planning and tenure security, services and homes are built for qualifying beneficiaries. The upgrading of a specific informal settlement takes place over a multi-year period, and is dependent on a number of factors including stipulated time periods required in the town planning process, applicable specialist studies including environment impact assessments, ecological, heritage, soil, vegetation and geotechnical studies. One of the major time delays are objections to the upgrading of informal settlements when they are situated adjacent to established townships.

(2) The National Housing Code: Upgrading of Informal Settlements Programme systematically details the process and procedure for the in situ upgrading of informal settlements in a structured manner. The key principles to be followed by implementers of the Programme are specified, including,

Community Engagement by their local authorities is of the utmost importance to ensure locally appropriate solutions. A feasibility study is to be conducted on the upgradeability of the settlement, and households must be profiled to determine beneficiaries. Detailed settlement level plans are to be development with the participation of the community;

Tenure: The Programme promotes security of tenure as the foundation for future individual and public investment. A check is done on land ownership to ensure security of tenure;

Suitable land: The programme will only provide funding in respect of informal settlements situated on land suitable for permanent residential development and within an approved IDP of the municipality concerned.

Service standards: The Programme provides funding for the installation of interim and permanent municipal engineering services. Where interim services are to be provided it must always be undertaken on the basis that such interim services constitute the first phase of the provision of permanent services. The nature and level of permanent engineering infrastructure must be the subject of engagement between the local authority and residents. Community needs must be balanced with community preferences, affordability indicators and sound engineering practice;

(3) (a) (i) (aa) During **2012-13** a total of **203** informal settlements were upgraded;

(bb) During **2013-14** a total of **113** informal settlements were upgraded;

(cc) During **2014-15** a total of **127** informal settlements were upgraded;

(dd) During **2015-16** a total of **95** informal settlements were upgraded;

(ii) Since 1 April 2016, **17** informal settlements were upgraded.

(b) to (d) We do report on project developments undertaken in our annual report and I would suggest that the Honourable member consult these reports for information dating back to 2012.

In addition, I wish to remind the Honourable member that during my budget Vote speech in the National Assembly on 3 May 2016, I indicated that urbanisation and the resultant mushrooming of informal settlements is something we are grappling with. I specifically said:

> *Last month we had the honour of hosting an International UN conference, in preparation for the Third UN Habitat Conference – a world conference that takes place every 20 years and which will now take place in Quito, Ecuador in October. We had 512 delegates from 54 different countries, and representatives of 54 governments, including 14 Ministers of Housing and we were given the opportunity to shape and influence the future of international human settlement discourse and subsequent policy and practice. The theme of the conference was "Urbanization and Informal Settlements". This was our choice as host country. We chose it because that is our present and pressing challenge with many of our people still living in squalor in places such as Khayelitsha, Gugulethu, Nyanga, Philippi, Soweto, Orange Farm, Polokwane, Mahikeng, Tshwane, eMlazi and all whom we dedicate today. The conference grappled with the staggering figures presented.*

Despite all the challenges we are faced with, South Africa is counted among the countries that have made significant contributions to improving the lives of those living in informal settlements, and we will continue to do so.

In the recent StatsSA survey, released in March, it is confirmed that amid growing urbanisation, the percentage of people living in informal settlements has dropped from 17% in 2002 to 11% in 2014.

Source: PMG, 2016f.